JOURNAL

of the

OPTICAL SOCIETY

of

AMERICA

MAY

1944

VOLUME 34 NUMBER 5

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA

THE JOURNAL OF THE OPTICAL SOCIETY OF AMERICA is the official organ of the Society, and is devoted to original papers on experimental, theoretical, and applied optics, in all its branches, including research, instruction, manufacture and design of optical instruments, as well as physiological and medical optics. It is published monthly by the American Institute of Physics Incorporated for, and under the editorial management and control of, the Optical Society of America, Prince and Lemon Streets, Lancaster, Pa.

NOTICE FOR CONTRIBUTORS

Manuscripts for publication should be sent to George R. Harrison, Massachusetts Institute of Technology, Cambridge 39, Massachusetts. Papers are published in English only. To be acceptable for publication manuscript should be original typewritten copy (not carbon copy) either double or triple spaced, with wide margins. References should appear as footnotes only and should be numbered consecutively to avoid repetition. References should include author's name, journal, volume number, page number, and year and should be listed in the order given here. Attention should be given to captions for tables and legends for figures. These in all cases should be complete in themselves so as to make the data intelligible to the reader without consulting the text of the article.

Special care should be given to mathematical expressions. The very simplest formulas only should be typewritten and all others carefully written in with pen and ink. Fractional exponents should be used to avoid root signs everywhere. Extra symbols should be used to avoid complicated exponents. The solidus (/) should be used wherever possible for fractions.

All illustrations should accompany manuscript and should always be referred to in the text. Line drawings must be made with India ink on plain white paper or on tracing cloth. Coordinate paper is not desirable, but if used must be blue-lined with all coordinates to be reproduced drawn with India ink. Lettering should be of sufficient size to be legible after reduction.

Proof and all correspondence concerning papers in the process of publication should be addressed to the Publications Manager, American Institute of Physics, 57 East 55 Street, New York 22, New York.

SUBSCRIPTION PRICE

United States and Possessions and Canada $7.00
Elsewhere $7.70

BACK NUMBERS

Complete set:
Vol. 1, 1917–Vol. 33, 1943—**$175.00**
Yearly rate: **$7.70**
Single copies: **$1.40** each 1917–1921;
$0.70 each thereafter.
Cumulative Index, 1917–1935: **$2.00**

Subscriptions, renewals, and orders for back numbers should be addressed to Prince and Lemon Streets, Lancaster, Pennsylvania, or to the American Institute of Physics, 57 East 55 Street, New York 22, New York.

The *Journal of the Optical Society of America* is published monthly at Prince and Lemon Streets, Lancaster, Pennsylvania.

Entered as second-class matter March 22, 1933, at the Post Office at Lancaster, Pennsylvania, under the Act of March 3, 1879. Acceptance for mailing at special rate of postage provided for in the Act of February 28, 1925, embodied in paragraph 4, section 538, P. L. and R., authorized March 22, 1933.

Journal of the

OPTICAL SOCIETY of AMERICA

VOLUME 34
No. 5

MAY
1944

Published monthly
at Prince and Lemon Sts., Lancaster, Pa.
by the
AMERICAN INSTITUTE OF PHYSICS
Incorporated
57 East 55 Street, New York 22, New York
FOR AND UNDER THE EDITORIAL SUPERVISION OF THE
OPTICAL SOCIETY OF AMERICA

THE OPTICAL SOCIETY OF AMERICA

THE OPTICAL SOCIETY OF AMERICA was organized in 1916 "*to increase and diffuse the knowledge of optics in all its branches, pure and applied, to promote the mutual interests of investigators of optical problems, of designers, manufacturers and users of optical instruments and apparatus of all kinds and to encourage cooperation among them.*" It aims to serve the interests of all who are engaged in any branch of optics from fundamental research to the manufacture of optical goods.

MEMBERSHIP

Regular Members: "Any person who has contributed materially to the advancement of optics is eligible for regular membership."

Associate Members: "Any person who is interested in optics is eligible for associate membership."

Corporate Members: "Any corporation desirous of encouraging the work of the Society is eligible for corporation membership."

Patrons: The payment, within any one year, of $1000 or more toward the expenses of the Society by any corporation or individual entitles such corporation or individual to the rank of patron.

The Society cordially invites to membership all who are interested in any branch of optics, either in research, in instruction, in optical or illuminating engineering, in the manufacture and distribution of optical goods of all kinds, or in physiological and medical optics.

ANNUAL DUES

Regular members, $8.50; Associate members, $6.00; Corporate members, $50.00. These dues include subscription to the Society's official journal, the *Journal of the Optical Society of America.*

CHANGES OF ADDRESS

All changes of address of members of the Optical Society of America should be communicated to the Secretary, Professor Arthur C. Hardy, Massachusetts Institute of Technology, Cambridge 39, Massachusetts.

Journal of the
OPTICAL SOCIETY of AMERICA

VOLUME 34, NUMBER 5 MAY, 1944

The Psychophysics of Color*

COMMITTEE ON COLORIMETRY

IN the development of a quantitative psychophysical science of color, the concepts of light and of color need be defined only by the procedures which are prescribed for their measurement. Such operational definitions are difficult to express concisely; consequently relational definitions are desirable in order to clarify the sense in which the words should be used. Such defini-itons can be formulated by stating the distinctions between these concepts, light and color, and the concepts related to them in the fields of psychology and physics. Light and color are psychophysical concepts according to both the relational and the operational definitions. These definitions are equivalent in significance, although the first type is most useful for guidance as to the correct use of the terms in general discussions and the second is essential for the precise definition of the concepts with which the measurements are concerned. The relational definitions are expressed in general terms, rather than in terms of quantitative concepts. The use of quantitative concepts would specialize the relational definitions in an undesirable manner.

1. DEFINITION OF LIGHT

The relations between radiant energy, light, and visual sensation can be defined in the following generalized and qualitative manner:

Light is the aspect of radiant energy of which a human observer is aware through the visual sensations which arise from the stimulation of the retina of the eye.

Light as thus defined is a psychophysical concept. Light is not identified with either radiant energy or with visual sensation. Light is one of the many conceivable aspects of radiant energy. The photographic effectiveness, the radiometric power, and the erythemal potency are examples of other aspects of radiant energy. Light, however, is the aspect of radiant energy of which human observers are aware through the immediate agency of the eye and the sensations and perceptions resulting from the stimulation of the retina. The concept of light is not limited to the quantitative aspect of radiant energy, because human observers are aware of more than just the quantitative aspect, which corresponds only to the brightness attribute of visual sensation. Light has many characteristics, only one of which is included in the specialized concepts of the quantity and flux of light. For instance, light may have, and usually has, the characteristics of non-uniform spatial distribution and of temporal fluctuations. In the following discussion reference to these characteristics will be made by use of the more concise phrase, "spatial and temporal inhomogeneities." In addition to these, light has characteristics which are referred to collectively as color. The quantitative characteristic of light is one of the characteristics included in the concept of color. Two other characteristics of light, which may be specified in terms of dominant wave-length and purity, are included also

* Chapter VI of the forthcoming Colorimetry Report.

TABLE I.

Physical Radiator (source of radiant energy) Radiation (process)				Psychophysical Luminator (source of luminous energy) Lumination (process)			
Radiometry	Symbol	c.g.s.	m.k.s.	Photometry	Symbol	c.g.s.	m.k.s.
Radiant energy	U	erg	joule	Luminous energy	Q	lumerg	talbot
Radiant density	u	erg/cm^3	joule/m^3	Luminous density	q	lumerg/cm^3	talbot/m^3
Radiant flux	P	erg/sec.	watt	Luminous flux	F	lumerg/sec.	lumen
Radiant emittance	W	erg/sec.$\times$cm^2	watt/m^2	Luminous emittance	L	lumerg/sec.$\times$cm^2	lumen/m^2
Radiant intensity	J	erg/sec.$\times\omega$	watt/ω	Luminous intensity	I	lumerg/sec.$\times\omega$	lumen/ω (candle)
Radiance	N	erg/sec.$\times\omega\times$cm^2	watt/$\omega\times$m^2	Luminance	B	lumerg/sec.$\times\omega\times$cm^2	lumen/$\omega\times$m^2 (candle/m^2)
Irradiance	H	erg/sec.$\times$cm^2	watt/m^2	Illuminance	E	lumerg/sec.$\times$cm^2	lumen/m^2 (lux)
Spectral reflectance	r			Luminous reflectance	R		
Spectral transmittance	t			Luminous transmittance	T		
Ratio of photometric quantity to corresponding radiometric quantity (standard units)=luminosity, K (luminous efficiency, lumens/watt) of radiant energy involved.							

Note: The nomenclature given in this table and used in this report differs in many details from the nomenclature recommended by the Illuminating Engineering Society, and adopted by the American Standards Association. These modifications of the standard nomenclature are proposed as somewhat simpler and more systematic, with the hope that they may be considered in a future revision of the standard nomenclature. Except for the omission of Greek symbols, Φ, ρ, τ, the symbols shown in this table are identical with those adopted by the American Standards Association. The standard symbols, ρ and τ for reflectance and transmittance do not indicate any distinction between the radiometric and the photometric concepts. The symbol ω used in Table I denotes a unit solid angle, the solid angle subtended by one square meter of surface of a sphere having a radius of one meter.

in the concept of color. Patterns of colors arise from non-uniform spatial distributions of light. Spatial non-uniformities and fluctuations of light result in the successive appearance of different colors and are not included among the characteristics which constitute color.

The evaluation of luminous flux originated before the precise evaluation of the other color characteristics of light was proposed. Consequently, the definition of this quantitative concept preceded, and was to a great extent independent of the definition of color. Therefore, although luminous flux is one of the characteristics of light used in defining color, it is possible, expedient, and, in fact, customary, to define luminous flux without reference to color. Luminous flux is the time rate of flow, emission, or incidence of light. In the sense required by this definition of luminous flux, light is the evaluation of radiant energy in terms of its capacity to produce the brightness attribute of visual sensation. This brightness-producing capacity of radiant energy is expressed in terms of luminous energy, in units of lumergs or talbots (see Table I). Various origins and manifestations of luminous flux are expressed in terms of luminance, illuminance, luminous emittance, or luminous intensity. The procedures prescribed in Chapter VII for the calculation of these quantities constitute the operational definition of light. Similar operational definitions of the other two characteristics of light, dominant wave-length, and purity, together with the operational definition of the quantity of light, constitute the operational definition of color.

2. DEFINITION OF COLOR

Color consists of the characteristics of light other than spatial and temporal inhomogeneities; *light being that aspect of radiant energy of which a human observer is aware through the visual sensations which arise from the stimulation of the retina of the eye.**

Color as thus defined is a psychophysical concept. Color is not identified with radiant energy nor is it identified with sensation. The characteristics of light which constitute color can be specified in terms of: (1) the appropriate photometric quantity, (2) dominant wave-length, and (3) purity. In a general way these characteristics of light correspond to the attributes of visual sensation: brightness, hue, and saturation. (Luminance corresponds most directly with brightness, yet other photometric concepts—illuminance, luminous intensity, and luminous flux—are also related to brightness under certain conditions.) Rarely, if ever, is light temporally constant or spatially uniform throughout the entire field of vision. Nevertheless, in accordance with this definition of color, the light from each portion of the field which is uniform and constant during an appreciable time has a definite color.

* This definition of color may be shortened by omitting everything following the semicolon (;), *provided that* the definition of light given in Section 1 is adhered to strictly. A note to this effect should always supplement the use of the short definition of color.

3. COLORS ASSIGNED TO OBJECTS

According to a strict interpretation of the definition of color given in the preceding section, it is not legitimate to attribute colors to objects, but only to the light reflected from them. Insofar as an object can be said to have a color, its color consists of the capacity of the object to modify the color of the light incident upon it. This capacity depends upon the variation of absorptance for various wave-lengths. When the absorptance varies with wave-length, the object is said to exhibit *selective absorption*, that is, the variation of absorptance with wave-length. In this sense, the color of an object may be considered to be a consequence of selective absorption, but color should not be considered synonymous with selective absorption. In addition to depending upon selective absorption, the color of an object depends upon the spectral distribution of the incident radiant energy, and *upon the psychophysical functions of human vision*.

There is no unique color of an object, since, among other things, the color depends upon the quality and intensity of the incident radiant energy. However, the color of the light reflected from an object in its most customary illumination (e.g., daylight or incandescent lamp light) is so familiar that it is usually considered to be the color typical of the object. In accordance with this practice, the luminous reflectance (or luminous transmittance) of the object, and the dominant wave-length and purity of the reflected light are considered to be characteristics of the object. These characteristics constitute the color assigned to the object.

Those varieties of light which are occasionally called "colorless" have a color according to the usage recommended here,† the purity of the color of neutral or "white" light being zero and the dominant wave-length being indeterminate. By extension of the concept that photometric quantity is one of the color characteristics of light, a dark object is considered to have a color different from light objects. This usage also eliminates the necessity for the phrase "light and color" which would be necessary if photometric magnitude were not one of the characteristics of light included in the concept of color. Photometry, the science of the measurement of the quantitative aspects of light, is thus included in colorimetry, the science of color measurement. It is not necessary to say that colorimetry is the science of the measurement of light and color, since the evaluation of the photometric magnitude, which is the measurement of one characteristic of light, is part of the measurement of color.

Colorimetry carries over the usage adopted in photometry by employing the concept of luminous reflectance (and luminous transmittance) as the most appropriate measure of light reflecting (or transmitting) efficiencies of objects. Luminous reflectance and transmittance, as well as dominant wave-length and purity of the colors attributed to objects, are dependent upon the spectral distribution of the incident energy. As stated in Chapter V, the term spectral distribution here refers only to the relative values of the energy per unit wave-length interval at every wave-length. The absolute values are referred to by the term spectral composition.

4. CHROMATICITY

The term chromaticity refers to the characteristics specified by dominant wave-length and purity. Any psychophysical specification of color which embodies the same information as domi-

† This recommendation is a retention of the usage which the Report of the Colorimetry Committee established in 1922 [see "Report of Committee on Colorimetry for 1920–21," J. Opt. Soc. Am. and Rev. Sci. Inst. 6, 527 (1922)]. A long footnote beginning on page 531 of that report explained in detail the reasons for which this usage was adopted after thorough and lengthy discussion within that committee. A summary of the objections to the opposite usage is found in the following quotation from the 1922 report. "On the other hand, if we define color in the restricted sense to exclude the gray series we find it necessary to exclude all considerations of brilliance (brightness) from the field of colorimetry. This means that if we are asked to specify the color of a gray object we must state that it has no color, and hence lies outside our province. Similarly, we should be compelled to affirm that certain browns are identical in color with certain yellows, oranges, and reds because they possess the same hue and saturation, although their brilliances (brightnesses) are quite different. The necessity for reactions of this sort on the part of the scientific colorimetrician would cause serious embarrassment in practice. It seems necessary to permit a certain degree of overlapping of the provinces of colorimetry and photometry, and possibly it would be desirable to include the latter under the former as a special branch." The word "brightness" has been inserted in parentheses after the word brilliance at two places in this quotation because this change of terminology is recommended by this report. The science of photometry has been included under the science of colorimetry as a special branch in this report, and the definitions of both light and color as psychophysical concepts are consistent with and, in fact, make possible this unification of, the two sciences.

nant wave-length and purity, but not photometric magnitude, is a specification of chromaticity. The measurement of color may be divided into two general problems: first, the measurement of luminance (or other appropriate photometric magnitude) and, second, the measurement of chromaticity. A diagram in which each point represents the chromaticity, independent of luminance of a color, is called a *chromaticity diagram*. This is a modification of the usage recommended in the 1922 report,[3**] in which the word chromaticity referred jointly to the subjective hue and saturation attributes of color sensation. The term *chromaticness* is now proposed for use with reference to these subjective attributes. This change is recommended because the need for and lack of a term referring to dominant wave-length and purity jointly has occasioned widespread usage of the word chromaticity in this sense. The substitution of "chromaticness" for use in the subjective sense is consistent with the reservation of the suffix "-ness" for terms descriptive of subjective phenomena. Thus, brightness and lightness have been defined in the subjective senses. A corresponding distinction between granularity, the objective concept, and graininess, a subjective attribute of non-uniform appearance, has been recommended elsewhere,[601] and a similar distinction between the objective concept, gloss, and the subjective attribute, glossiness, has also been proposed.[602] The new name for the photometric concept, luminance, in place of the long accepted term, brightness, is strongly recommended[10] both for the sake of uniformity of nomenclature outlined above and because conflicting popular usage of the word brightness is the origin of frequent misunderstandings in public discussions of photometric problems. The word, brilliance, recommended in the 1922 report[3] as the name for one attribute of visual sensation, has been abandoned because of frequent association of other concepts with this term.

5. LUMINOSITY

The measurement of luminance (and other photometric quantities which are related to the brightness attribute of visual sensation) is based on the acceptance of certain data which are frequently called "visibility" data. The term "luminosity"[603, 523†] has superseded the word "visibility" by action of all of the authoritative bodies which were responsible for the adoption of the latter term. These include the Illuminating Engineering Society,[13] the International Commission on Illumination, and the Advisory Committee on Photometry of the International Committee on Weights and Measures. The decisions of the two international bodies were made at meetings in 1939 and, consequently, formal publication has been delayed on account of the war.

It is evident from visual examinations of the spectrum of any source (such as the sun), which radiates approximately equal amounts of energy per unit wave-length interval throughout the visible spectrum, that equal amounts of energy of different wave-lengths do not produce visual sensations having equal brightnesses. A luminosity curve shows as ordinates the capacities (reciprocals of the required radiance) of radiant energy of the various wave-lengths to evoke for a particular observer visual sensations of equal brightness. Although the luminosity curve may be considered to represent the response of the visual receptors to spectral stimuli, this response cannot be measured directly in a manner analogous to the measurement of the response of a photoelectric cell. Some indirect method must be employed for the determination of the luminosity curve.

6. DETERMINATION OF LUMINOSITY DATA

The standard luminosity data represent the concordant results of several methods of heterochromatic photometry. The cascade, or step-by-step method of heterochromatic photometry employs a photometric field divided into two adjacent areas which can be irradiated with radiant energy from separate sources. One part of the field may be irradiated so as to produce a known radiance of homogeneous radiant energy of known wave-length. The other part of the field may be irradiated with homogeneous radi-

** References numbered from 1 to 16 are given in the first article of this series, J. Opt. Soc. Am. **33**, 534 (1943).
[601] E. M. Lowry, J. Opt. Soc. Am. **26**, 65 (1936).
[602] L. A. Jones, J. Opt. Soc. Am. **6**, 146 (1922).
[603] Anonymous, Rev. Sci. Inst. **7**, 322 (1936).
† References 501 to 573 are given in the third article of this series, J. Opt. Soc. Am. **33**, 183 (1944).

ant energy having a slightly different wave-length. The radiance in the second half of the field may be adjusted until the visual difference in appearance between the two part-fields is minimized. The visual difference in appearance can be completely eliminated only if the difference between the two wave-lengths is made sufficiently small. Usually it is not necessary to use wave-lengths so close together, and frequently it is inconvenient to use such small wave-length intervals. The reciprocal of the ratio of the radiance necessary for the minimization or elimination of visual difference between the two fields is the ratio of the ordinates of the luminosity curve for the two wave-lengths. The cascade method consists of comparing in this manner one wave-length with a second neighboring wave-length, then comparing the second with a third, then the third with a fourth, progressing in this way throughout the spectrum until all the ordinates of the relative luminosity curve have been determined for the observer.

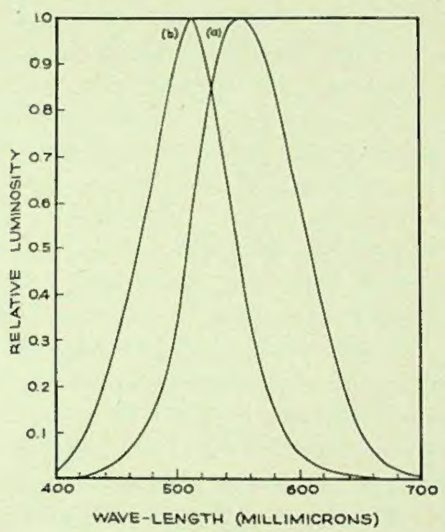


FIG. 1. Luminosity curves: (*a*) Standard I.C.I. luminosity curve (reference 605) for photopic vision (observer adapted to high luminance); (*b*) Weaver's (reference 606) average curve for scotopic vision (observer adapted to low luminance).

Other methods, notably flicker photometry, may also be used for the determination of luminosity data. In the flicker photometer method two homogeneous radiant energies, which may be widely different in wave-length, are presented alternately to the eye at such a rate that, although the differences of chromaticity cannot be detected, the brightness differences can be appreciated as a flicker.[604] When the radiances have been adjusted to eliminate (or minimize) the flicker, their ratio is the reciprocal of the ratio of the corresponding ordinates of the luminosity curve. Experiments with the same observer, employing certain standard conditions of observation in both methods, demonstrated that under certain conditions the step-by-step method and the flicker photometer method yield concordant luminosity data.[604]

7. DEPENDENCE OF LUMINOSITY DATA ON CONDITIONS OF OBSERVATION

The curve obtained by the use of either of these methods applies only for the conditions of observation employed during the measurements. Curve (*a*) in Fig. 1[605, 606] shows the standard luminosity data adopted[605] by the International Commission on Illumination in 1924. These data were adopted after a study of the results of Gibson and Tyndall,[607] employing 52 observers, confirmed the results of similar measurements by Coblentz and Emerson[608] and others. These luminosity data are representative of normal vision under good lighting conditions and can be recommended for general use.[609]

Curve (*b*) in Fig. 1 shows the average luminosity curve obtained by Weaver,[606] for 14 observers adapted to a test field illuminance of 0.000065 footlambert, representative of scotopic vision such as obtained by star light or in the subdued lighting of photographic darkrooms.

The comparison of curves (*a*) and (*b*) indicates the type and degree of variation of the luminosity curve which can be expected to result from changes in the conditions of observation. Intermediate adaptive states yield curves intermediate between the two shown, and higher intensities may shift the maximum of the curve to wave-

[604] H. E. Ives, Phil. Mag. [6] 24, 149, 352, 744, 845 (1912).

[605] Proceedings of the Sixth Session International Commission on Illumination (Geneva, 1924), p. 67.

[606] K. S. Weaver, J. Opt. Soc. Am. 27, 26 (1937).

[607] K. S. Gibson and E. P. T. Tyndall, Sci. Pap. Bur. Stand. 19, 131 (1923–24); Trans. Illum. Eng. Soc. 19, 176 (1924).

[608] W. W. Coblentz and W. B. Emerson, Bull. Bur. Stand. 14, 167 (1918–19).

[609] K. S. Gibson, J. Opt. Soc. Am. 30, 51 (1940).

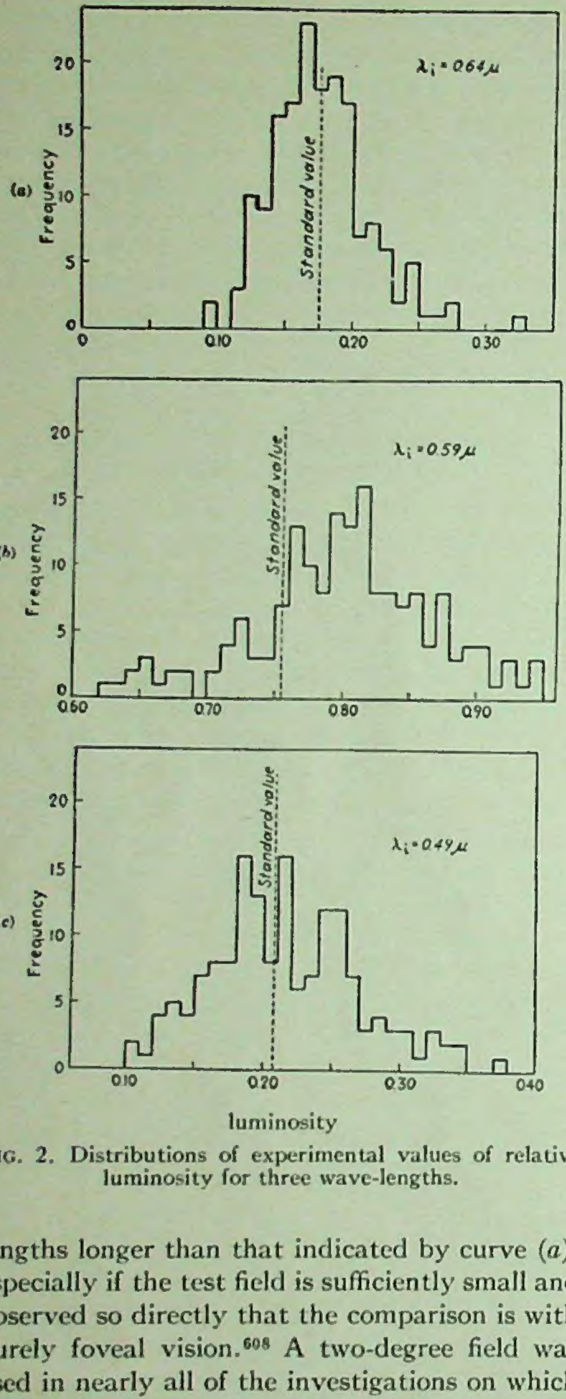


FIG. 2. Distributions of experimental values of relative luminosity for three wave-lengths.

lengths longer than that indicated by curve (*a*), especially if the test field is sufficiently small and observed so directly that the comparison is with purely foveal vision.[608] A two-degree field was used in nearly all of the investigations on which the standard curve was based. Since the use of several luminosity curves would be a source of endless confusion and ambiguity in photometric measurement and specification, curve (*a*) is recommended for use in connection with all problems involving customary lighting conditions. For the purposes of standardization and interlaboratory comparison of photometric data, the use of standard luminosity data is essential.

8. INDIVIDUAL VARIATIONS OF LUMINOSITY DATA

Even with carefully chosen and controlled conditions of observation, the luminosity curve obtained for each observer differs more or less from the curves of other observers. Except for that small but not negligible part of the population whose color vision can be classified as anomalous, most people are remarkably similar with respect to their abilities to see and to distinguish colors. It might be expected that all such individuals would require the same relative radiances in luminosity determinations and that they would therefore have the same luminosity curve. The results of a large number of such measurements do not fulfill this expectation. Figure 2* shows the distribution of the results of measurements made by Coblentz and Emerson[608] for 125 observers, and by Gibson and Tyndall[607] for 52 observers. This figure demonstrates the wide range of value of luminosity obtained in experiments with "normal" observers. The abscissas are values of the luminosity of the indicated wave-length obtained with different observers, while the ordinates represent the numbers of observers obtaining these values. At 490 mμ, for instance, luminosity values ranging from 0.10 to 0.38 were obtained, though the greatest number of observers obtained values in the neighborhood of 0.20. Similar results were obtained at other wave-lengths, as is illustrated by the distribution of results shown for 590 mμ and 640 mμ. An observer chosen at random may give a value anywhere in the range indicated by the distribution curves such as are shown in Fig. 2, although his value is more likely to be near the average value than at the extremes of the range.

9. STANDARD LUMINOSITY DATA

On the basis of a comparison of their results with the results obtained by previous investigators, Gibson and Tyndall recommended for adoption a particular set of luminosity values. These values were adopted by international

* Reprinted from *Scientific Basis of Illuminating Engineering*, P. Moon, by permission of McGraw-Hill Book Company, New York.

agreement in 1924 at the Geneva meeting of the International Commission on Illumination.[605] The adoption of these data was an arbitrary procedure, since somewhat different data would have been obtained with another group of observers, and different conditions of observation would have resulted in different data even with the same group of observers. No one curve can represent the response of any eye under all conditions of vision. No average of a finite number of luminosity curves can be proved to be the exact average for the whole population. The particular data which have been standardized probably differ less seriously from the data for an observer selected at random than would any other arbitrarily selected set of data. The importance of the standard luminosity data is that these particular values have been accepted internationally and may be used to establish the photometric quantities by definition. Thus defined, these photometric quantities have meaning even when the conditions of observation are not standard. The direct correlation of these photometric quantities with the brightness attribute of visual sensation is not assured unless the standard conditions of observation are fulfilled and when the observer has a luminosity curve indistinguishable under these conditions from the standard luminosity curve. The utility of these quantities, even when any or all of these conditions are not fulfilled, lies in the fact that the adoption of a standard luminosity curve for general use is the most convenient basis of general agreement concerning photometric concepts. In practice, it is customary not to limit the use of the standard luminosity data to problems involving photopic (daylight and normal reading) conditions, though it is probable that we should look forward to the ultimate establishment by international agreement of standard luminosity data for scotopic (extremely low) luminances and for the intermediate (mesopic) luminances corresponding to a transition from photopic to scotopic vision. It is well known that application of photopic luminosity data to the scotopic state of adaptation gives results very noticeably inconsistent with visual observations, and for some problems, such as the photometry of luminescent and phosphorescent materials, supplementary alternate procedures are being devised. The complications introduced by these alternate procedures may, however, be expected to limit their use. For example, none of the computational methods of illuminating engineering which are based on the inverse square law of diminution of illuminance with distance can be used if special luminosity data for scotopic and transitional condition be adopted. This law assumes by implication that photometric quantities are related to radiometric quantities through one invariant luminosity function, valid for all conditions. The ordinary photometric concepts of luminous flux, emittance, intensity, reflectance, and transmittance would become useless and meaningless if the luminosity of radiant energy were defined as dependent on the luminance of the object observed. For most of the problems of illumination and colorimetry the advantage of avoiding inconsistency with relative luminosities which are observed by scotopic and mesopic vision is negligible compared to the serious complications which would be introduced by the adoption of multiple luminosity standards for general use. We may therefore confidently expect that such standards will be restricted to special problems.

10. LUMINOSITY: RELATIVE LUMINOUS EFFICIENCY

As curve (*a*) in Fig. 1 is drawn, it indicates only the relative values of luminosity. It is, however, quite possible to give these ordinates a more definite quantitative significance. Assuming that capacity to produce brightness is expressed quantitatively in terms of some suitable unit and that the radiant power possessing this capacity is also expressed in a suitable unit, it is evident that the ratio of these two quantities is an expression of the brightness-producing capacity per unit of radiant power. This is the usual form of the expression of efficiency. We may, therefore, say that the ordinates of curve (*a*) in Fig. 1 are proportional to the efficiency of radiant power for the production of brightness. This efficiency will be called the ***luminous efficiency*** of radiant energy. Energy is not the most suitable quantity with which to discuss the correlation with brightness. The brightness produced depends not upon the amount of energy incident upon the retina, but rather upon the time rate

at which it is incident upon a unit area of the retina. Thus brightness is determined by the irradiance of the retina. The standard metric unit of irradiance is a watt per square meter.

In attempting to choose a unit for the expression of the brightness-producing capacity of radiant energy, we encounter the difficulty of measuring sensation. Whether or not it is possible to measure sensation, the fact remains that up to the present time no methods or unit for the purpose have been generally accepted. We are forced, therefore, to turn to the psychophysical correlate of brightness, which is retinal illuminance. The standard metric unit of illuminance is a lumen per square meter. This unit has been named the *lux*. The luminous efficiency of radiant energy can therefore be expressed in terms of lumens per watt, the units of area common to the unit for retinal illuminance and irradiance cancelling when the ratio is taken.

11. ABSOLUTE LUMINOUS EFFICIENCY OF RADIANT ENERGY

The photometric and radiometric units of flux, lumens and watts, have been standardized independently of each other. The standardization of the relative luminosity data represented by curve (*a*) in Fig. 1 permits the calculation of the most probable value†[610] of the maximum luminous efficiency of radiant flux, 650 international lumens per watt of radiant flux having the wave-length 555 mμ. If each of the ordinates of the standard luminosity curve is multiplied by 650, the resulting values are the luminous efficiencies of radiant flux of the corresponding wave-lengths. The curve representing these results may be called the absolute luminosity curve for radiant flux. The ordinates are lumens per watt, with the maximum value of 650 international lumens per watt at the wave-length 555 mμ.

Both the standard luminosity curve and the absolute luminosity curve are commonly interpreted in either of the following two manners:

(1) Descriptive of a property of the eye, they both indicate the spectral distribution of retinal brightness sensitivity.

(2) Descriptive of a property of radiant flux, they specify in relative and absolute terms, respectively, the spectral distribution of luminous efficiency.

It is the purpose of this report to emphasize that the luminosity function, and also the similar colorimetric functions, are not properly described in either of the above manners, but that they are psychophysical functions, medial in nature between psychological functions (as implied by 1) and physical functions (as implied by 2).

It is recommended that luminosity be defined as the equivalent of the phrase, luminous efficiency of radiant energy, in either relative or absolute values according to the context. The standard symbol for the absolute values of luminous efficiency will be K. The standard symbol for the relative values will be $\bar{y}$. The absolute luminosity of homogeneous radiant energy of wave-length λ may be designated by K_λ, the relative value by $\bar{y}_\lambda$.

12. LUMINOUS FLUX

When radiant flux is evaluated with respect to its capacity to evoke the brightness attribute of visual sensation it is called luminous flux, and

† This value (650 international lumens per watt) has been adopted by the Committee on Nomenclature and Standards of the Illuminating Engineering Society of America (reference 13), and is consistent with values calculated by H. T. Wensel (reference 610). It is based upon: (1) the standard values of relative luminosity adopted by the International Commission on Illumination in 1924; (2) the value, 1.436, of the constant, c_2 in Planck's formula for radiant flux from a complete radiator; and (3) the luminance of a complete radiator at the freezing point of platinum, which on the basis of measurement made in the national laboratories of the United States, England, France, and Germany is generally considered (reference 555) to be 58.9 international candles per square centimeter. The value of the constant, $c_2 = 1.432$, adopted in 1927 for the purpose of establishing the international temperature scale is no longer tenable on the basis of either direct experimental determinations or calculations from accepted values of the atomic constants involved (reference 610). Therefore, the value for the maximum luminous efficiency of radiant flux calculated from the value of $c_2 = 1.432$ cannot be regarded as the most probable value of this conversion factor. The value so computed is 625 international lumens per watt, in approximate agreement with values which have been published in the past. A new system of photometric units has been proposed (references 610 and 555), in terms of which the luminance of a complete radiator at the temperature of freezing platinum would be 60 "new" candles per square centimeter. If and when such units are put into effect, the most probable value of the maximum luminous efficiency of radiant flux will be 663 "new" lumens per watt. Birge (reference 521) has recommended 1.4385 as the most probable value of c_2. This value implies 671.3 international lumens per watt as the maximum luminous efficiency. The lower value 650 is used in this report, consistent with Wensel's value of $c_2 = 1.436$, which was used in the preparation of extensive tables in Chapter VII.

[610] H. T. Wensel, J. Research Nat. Bur. Stand. **22**, 375 (1939).

this capacity is expressed in lumens. Curves showing the spectral distribution of luminous flux from various sources have frequently been called the luminosity curves of the sources. This use of the word luminosity is inconsistent with the definition of luminosity now recommended. The ratio of the total luminous flux to the total radiant flux from a source (expressed in lumens per watt) is the luminosity K, or luminous efficiency of the radiant flux from that source.

13. LUMINOUS ENERGY

The concept of luminous energy,[10] which corresponds to radiant energy, is rarely used explicitly although it is the most appropriate measure of the economic value of light. The c.g.s. unit of luminous energy is the lumerg.[10] One erg of radiant energy having a luminous efficiency of K lumens per watt constitutes K lumergs of luminous energy. The m.k.s. unit of luminous energy, the talbot,[10] is ten million lumergs. One joule ($=10^7$ ergs) of radiant energy having a luminous efficiency of K lumens per watt provides K talbots of luminous energy, and a luminous flux of one talbot per second is one lumen.

14. LUMINOUS DENSITY

Luminous density[10] is the luminous energy contained in a unit volume of space, analogous to radiant density, and may be used in discussions of illumination problems in the same manner as radiant density is used. The c.g.s. unit of luminous density is the lumerg per cubic centimeter. The m.k.s. unit is the talbot per cubic meter. The luminous density is K times the radiant density expressed in the corresponding units.

15. LUMINOUS INTENSITY

Luminous intensity is expressed in candles. A luminous intensity of one candle produces one lumen of luminous flux through an area subtending a solid angle of one steradian measured from the source. The luminous intensity of a source is K times the radiant intensity expressed in watts per steradian. The concept of luminous intensity, like radiant intensity, applies only to sources that are negligibly small compared to the distance traveled by the emitted flux.

16. LUMINANCE

Luminance is luminous flux (lumens) per unit solid angle emitted per unit projected area of the source. A perfectly diffusing surface is one for which luminance is independent of direction of observation. The intensity (lumens per unit solid angle, or candles) of a perfectly diffusing plane source is therefore proportional to the cosine of the angle between the perpendicular to the surface of the source and the direction of observation. A convenient unit of luminance is the candle per square centimeter of projected area. The primary standard of the photometric system of units is the luminance of the surface of a complete radiator at the temperature of freezing platinum.* This standard[555] has a luminance of 60 candles per square centimeter. As a consequence of the fundamental physical characteristics of this standard, the photometric quantities can be determined from the corresponding radiometric quantities. Table I indicates these correspondences.[10] The luminance of any surface is the product of the radiance of that surface by the absolute luminosity K of the energy radiated by the surface. If the radiance is expressed in watts per steradian per square centimeter of projected area, K times the radiance will be the luminance expressed in candles per square centimeter of projected area of lumens per steradian per square centimeter of projected area. Many other units are in use for the specification of luminance, and the following table gives their equivalents in candles per square centimeter (projected areas are involved in all instances):

1 candle per square meter
$= 10^{-4}$ candle per square centimeter;
1 candle per square foot
$= 0.0010764$ candle per square centimeter;
1 lambert
$= 1/\pi$ candle per square centimeter;
1 millilambert
$= 1/1000\pi$ candle per square centimeter;
1 footlambert
$= 0.0003426$ candle per square centimeter.

* This standard is not yet in use, and its introduction in practice must await the exchange of calibrated working standards among the various national laboratories.

17. LUMINOUS EMITTANCE

The luminous emittance of a surface is the total luminous flux emitted per unit area of the surface. A convenient unit of luminous emittance is the lumen per square centimeter. When the luminance, $B(\theta)$, expressed in candles per square centimeter, varies with the angle of observation θ, measured from the normal to the surface, then the luminous emittance is given by the formula

$$L = 2\pi \int_0^{\pi/2} B(\theta) \sin\theta \cos\theta d\theta.$$

If the luminance of the surface is the same in all directions, the luminous emittance is found by integration to be π times the luminance expressed in terms of the candle per square centimeter. The luminous emittance of such a perfect diffuser is exactly equal to the luminance expressed in lamberts. If the luminance of such a surface is expressed in footlamberts, the luminous emittance is an equal number of lumens per square foot.

18. ILLUMINANCE

Illuminance is the luminous flux incident per unit area of a surface. Illuminance is the product of the irradiance of a surface and the luminosity K of the energy with which the surface is irradiated. This quantity has been commonly called "illumination" in the past.† The m.k.s. unit of illuminance, named the *lux*, is equal to one lumen per square meter. This unit is also known as the "meter-candle." The "footcandle" is the English unit of illuminance and is equal numerically to one lumen incident per square foot.

19. RETINAL ILLUMINANCE

Retinal illuminance has already been mentioned as the psychophysical correlate of brightness. The irradiance of the fovea of the schematic eye specified by Helmholtz is 0.3267 times the product of the square of the diameter (in centimeters) of the entrance pupil of the eye and the radiance of the object imaged on the fovea. Substitution of the photometric equivalents for the radiometric quantities involved reveals that, ignoring losses by reflection and absorption, the foveal illuminance (in lux) is 1.04 times the product of the square of the diameter (in centimeters) of the entrance pupil of the schematic eye and the luminance (in millilamberts) of the object which is focused on the fovea. Since four percent is a fair estimate of the loss by reflection at the cornea, the retinal illuminance expressed in lux can be taken as the product of the transmittance of the ocular media times the square of the diameter of the pupil (in centimeters) and the luminance (in millilamberts) of the object viewed.

A special unit of retinal illuminance, the photon, was introduced by Troland[611] in 1917 and has been employed to some extent, but its use is circumscribed by the danger of confusion with the elementary quantum of radiant energy which is generally called the photon. The renaming of the unit of retinal illumination as the troland, in honor of Dr. L. T. Troland, its originator and chairman of the Committee on Colorimetry of the Optical Society of America in 1921–22, is recommended. In accordance with this suggestion the troland is defined as the retinal illuminance produced by luminance of one candle per square meter when the apparent area of the entrance pupil of the eye is one square millimeter. Data already expressed in photons should be expressed either in trolands, without the necessity for any numerical conversions, or they should be converted to lumens per square meter (lux) by multiplying the number of trolands by 0.00400 and by the transmittance of the ocular media. According to recent measurements,[612] the transmittance of the media of the living human eye does not exceed 0.50 and decreases markedly with advancing age.

20. LUMINOUS REFLECTANCE

The concepts of luminous emittance and luminance are applicable to self-luminous surfaces and also to surfaces which are made luminous by diffuse reflection* or transmission of luminous flux which is incident upon them. The following

† In accordance with the recommendation (reference 10) that the suffix "-tion" be reserved for names of processes, the terms, illumination, reflection, and transmission, have been used as the names of processes in which radiant energy and light are involved. These terms will not be used to refer to quantities, nor to entities for which quantitative evaluation is conceivable.

[611] L. T. Troland, J. Exper. Psychol. 2, 1 (1917).

[612] E. Ludvigh and E. F. McCarthy, Arch. Ophth. 20, 37 (1938).

* See dagger footnote, left-hand column, p. 254.

discussion is based on the assumption that the illuminance, luminous emittance, and luminance are all expressed in units involving the same unit of area. Convenient metric units are the lumen per square centimeter (both illuminance and emittance) and the lambert. The corresponding English units are the equivalent footcandle, the lumen per square foot, and the footlambert. The luminous emittance of a reflecting surface is proportional to illuminance. When any consistent set of units is used the factor of proportionality for any surface, that is, the ratio of the luminous emittance to the illuminance of a reflecting surface, is its luminous reflectance. The luminous reflectance of a surface depends not only upon the nature of the surface but also upon the spectral distribution of the incident radiant energy and the relative amounts of the energy which are incident upon the surface from various directions. These conditions must therefore be controlled in every measurement of luminous reflectance, and the specification of luminous reflectance should be supplemented with an unambiguous statement of these conditions of measurement. The luminous reflectance of a surface for light consisting of radiant energy of a single wave-length is equal to the radiant reflectance of the surface for that wave-length, and for the same spatial conditions of irradiation. If the incident radiant energy is a heterogeneous mixture (i.e., consisting of two or more wave-lengths) the proportions of which are the same for all directions of incidence, the luminous reflectance is given by the ratio

$$R=\int_0^\infty r_\lambda K_\lambda H_\lambda d\lambda \Big/ \int_0^\infty K_\lambda H_\lambda d\lambda.$$

In this formula

$H_\lambda d\lambda$ = irradiance of surface with energy in wave-length band extending from $\lambda-\frac{1}{2}d\lambda$ to $\lambda+\frac{1}{2}d\lambda$.

K_λ = luminosity of radiant energy of wave-length λ.

r_λ = reflectance of surface for wave-length λ, for specified angular conditions of irradiation.

R = luminous reflectance of surface irradiated with spectral distribution H_λ for same angular conditions of irradiation for which the values r_λ were measured.

Since luminous reflectance is the ratio of two integrals, in both of which the product $K_\lambda H_\lambda$ appears, this product may be expressed in any units. It is convenient to choose units for tabulated values of this product such that the denominator of the expression for R equals unity:

$$\int_0^\infty K_\lambda H_\lambda d\lambda = 1.0.$$

The luminous reflectance of a perfectly diffusing reflecting surface is independent of the spatial conditions of illumination. For such surfaces the luminance is also independent of the direction from which the surface is observed. The luminance of a perfectly diffusing reflecting surface is equal to its luminous reflectance multiplied by the luminance of a perfectly reflecting, perfectly diffusing surface illuminated in the same manner. When expressed in lamberts, the luminance of a perfectly diffusing surface is numerically equal to the luminous emittance expressed in lumens per square centimeter. Therefore, the luminance (in lamberts) of such a surface is equal to the product of the illuminance (expressed in lumens per square centimeter) by the luminous reflectance. In the English system the luminance (in footlamberts) of a perfectly diffusing surface is equal to the product of the illuminance (footcandles) by the luminous reflectance. The luminance expressed in footlamberts is π times the candles per square foot.

21. DIRECTIONAL LUMINOUS REFLECTANCE

The luminance of an imperfectly diffusing surface may be compared with the luminance of a perfectly reflecting, perfectly diffusing surface which is similarly illuminated. The ratio of these luminances is a quantity which frequently differs from luminous reflectance. This ratio of luminances will be called directional† luminous reflectance. The corresponding ratio of radiances of

† The name apparent reflectance has sometimes been applied to the quantity designated here by directional reflectance. The adjective directional is preferred to apparent in this connection because the latter seems to convey undesirable impressions of uncertainty and indefiniteness, and even some suggestion of the subjective attributes of visual perception. The adjective directional is appropriate in that the direction of observation must be specified as well as the conditions of illumination, in order to give the quantity a specific meaning.

the surfaces, for homogeneous incident radiant energy and for the same spatial conditions of irradiation and observation, may be called directional spectral reflectance.

The directional luminous reflectance of a surface for light consisting of radiant energy of a single wave-length is equal to the directional spectral reflectance of the surface for that wave-length, and for the same spatial conditions of irradiation and observation. The directional luminous reflectance of a surface irradiated with a mixture of energies of two or more wave-lengths is given by the formula written above for luminous reflectance. In the calculation of directional luminous reflectance, however, the directional spectral reflectances measured for the appropriate conditions of irradiation and observation should be substituted for spectral reflectance r_λ in the formula. The luminance (in lamberts) of an imperfectly diffusing surface is numerically equal to the product of the illuminance of the surface by the directional luminous reflectance. The directional reflectances, and therefore also the luminance, vary with the direction from which the surface is observed, as well as with variations in spatial distribution of the incident light.

22. SPECULAR LUMINOUS REFLECTANCE

Directional luminous reflectance is not convenient for the specification of the reflectance of mirrors because of the inconveniently high values of this quantity measured in the direction of specular reflection of mirrors. Mirror surfaces are more conveniently specified by ***specular luminous reflectance***, which is the ratio of the luminance of an object as observed by reflection in a mirror to the luminance of the object viewed directly. Specular reflectance can be defined for any wave-length as the ratio of the radiances measured by reflection and directly. Specular reflectance can be substituted in the formula given above, and the specular luminous reflectance of the surface can be calculated for any spectral distribution H_λ of the energy radiated by the object which is observed by specular reflection. Specular reflectances usually vary with the angle of incidence and reflection, and, except in the case of non-selective mirrors, the specular reflectances also vary with the spectral distribution of the incident energy. In most cases specular reflection is accompanied to an appreciable extent by diffuse reflection. This diffusion can be specified in terms of directional luminous reflectances for various angles of incidence and observation. With high quality mirrors, practically no diffuse reflection occurs, and in this case there is no necessity for distinguishing between specular reflectance and the concept of reflectance which was defined as the ratio of emitted to incident flux per unit area.

23. LUMINOUS TRANSMITTANCE

Objects which transmit light without appreciable scatter are said to be non-diffusing. Objects which transmit light with appreciable diffusion are called diffusing. The luminous transmittance of any object is the ratio of the luminous flux transmitted by the object to the luminous flux incident. In the case of an object having uniform thickness, the luminous transmittance is also equal to the illuminance of the externally illuminated side divided into the luminous emittance of the opposite side. In general, the luminous transmittance of an object is dependent upon both the spectral and the spatial distributions of the incident radiant energy. The luminous transmittance of an object for radiant energy of a single wave-length is equal to the transmittance t_λ for that wave-length and for the same spatial conditions of irradiation. The luminous transmittance of a non-fluorescent object is given by the formula

$$T=\int_0^\infty t_\lambda H_\lambda K_\lambda d\lambda \bigg/ \int_0^\infty H_\lambda K_\lambda d\lambda.$$

As in the calculation of luminous reflectance R the luminous transmittance T will be simply the value of the numerator of this expression if the values of the product $H_\lambda K_\lambda$ used are expressed in units such that

$$\int_0^\infty H_\lambda K_\lambda d\lambda = 1.0.$$

24. DIRECTIONAL LUMINOUS TRANSMITTANCE

The luminance of a diffusely transmitting object is in general dependent upon the direction from which the object is viewed. A quantity analogous to directional reflectance may be useful for the specifications of such objects. This quan-

tity may be called directional luminous transmittance, and may be defined as the ratio of the luminance (candle per unit area) of the second surface to the illuminance (lumen per unit area) of the first surface. The spatial distribution of the incident light, and of the light collected for the determination of the luminance, should supplement the specification of directional luminous transmittance. The calculation of directional luminous transmittance from spectrophotometric data would require the definition and measurement of analogous quantities, which might be named directional radiant transmittances. The concept of directional transmittance is obviously inappropriate for the discussion of the light transmitted by practically non-diffusing (transparent) objects. Luminous transmittance is suitable for the specification of such objects. A concept named regular transmittance may be useful for the specification of the transmittance for image-forming flux by partially diffusing materials. The concept of directional transmittance would then be adequate for the specification of the light transmitted diffusely by such materials.

25. MEASUREMENT OF CHROMATICITY

In photometry chromaticity differences between samples and standards are sources of difficulty and error. In general, chromaticity differences between samples and standard cannot be eliminated in photometric measurements. The various methods of heterochromatic photometry are designed to minimize the noticeability of chromaticity differences, without interfering with the determination of the relative photometric magnitudes. The measurement of chromaticity, on the other hand, is based on the adjustment of the chromaticity as well as the photometric magnitude of the standard until all visually apparent differences between the sample and standard are eliminated. The most accurate methods of colorimetry are those in which the noticeability of any residual chromaticity difference is greatest, and in which the physical specifications of the adjustable standard are determinable with the greatest accuracy. Chromaticity may be specified in terms of any optical arrangement which provides for the adjustment of the chromaticity of the standard. Specifications of chromaticity in terms of any one arrangement can be translated into terms of any other system of colorimetry by use of conversions based on the fundamental data which will be described. Dominant wave-length and purity are specifications of chromaticity in terms of a method of colorimetry which has been used very rarely in recent years. However, the results of more reliable, modern methods of colorimetry are very frequently converted into dominant wave-length and purity, and these specifications are usually preferred for the specification of chromaticity because colors are more readily visualized when specified in these terms. The dominant wave-length of a sample is the wave-length of spectrally homogeneous radiant energy which would have to be mixed with an appropriate amount of achromatic, or "white," light in order to match the chromaticity of the sample. The purity is an expression of the proportion of the spectrally pure component in the mixture matching the chromaticity of the sample. When the amounts of the components and the mixture are expressed in terms of luminance, this proportion is called colorimetric purity. Dominant wave-length and purity are rarely determined directly, but are usually computed from the results of other methods of colorimetry, such as the method of three-color synthesis which will be described.

26. THREE-COLOR METHOD OF COLORIMETRY

In the three-color method of colorimetry, each sample is matched with a mixture of three components of light having different chromaticities. The amounts (for example, luminance) of these components are independently variable. The chromaticities of the components are not variable, but may be selected initially so as to make possible the synthesis of a suitably large gamut of colors. The components employed in this system, called trichromatic colorimetry, are usually such that they can be described as red, green, and blue. The amounts of three components, of which the chromaticities are also stated, constitute a specification of the color synthesized by the mixture of these three components. The chromaticity of the synthesized color may be specified in terms of the proportion of each of two of the components in the trichromatic mixture. These proportions are called *trichromatic coefficients*. Most samples can

be matched by a suitable mixture of the three components. In this case, the amounts of the components are recorded as positive quantities. However, some samples of light cannot be matched with such a mixture of the three components. In most cases of this kind, it is possible to establish a match between a mixture of the sample light with one of the components and a mixture of the other two components. In this case, the amount of the component which is mixed with the sample light is considered to be a negative quantity. This quantity is subtracted from the sum of the amounts of the two components of the other mixture involved in the color match. The coefficients (proportions) of the components, which specify the chromaticity of the sample, are computed on the basis of this net total. The coefficient of the component which is mixed with the sample light is considered as a negative quantity, so that the algebraic or net total of the coefficients of the components necessary to establish the match is unity in all cases.

27. THREE-COLOR MIXTURE AND MATCHING

A series of examples will serve to clarify the procedures outlined above. A sample of "white" light such as daylight may be visually compared with a mixture of the three components (usually red, green, and blue). Any observer will find it possible to adjust the amounts of these three components in the mixture until it has exactly the same visual appearance as the sample of daylight. This procedure of adjusting the amounts until all visually apparent differences are eliminated is called color matching. The amounts of the three components of the mixture are recorded as positive quantities. These three quantities can be added together and the total divided into each individual amount in order to compute the coefficient of each component required for the color match. These coefficients constitute a specification of the chromaticity of the sample (daylight) in terms of the three components and for the observer who established the match. It is obvious that since the sum of the three coefficients is always unity, only two of them are necessary for the specification of the chromaticity.

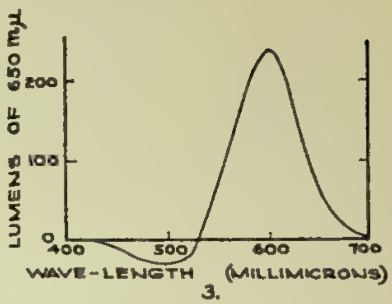
If the daylight is passed through a yellow glass before the match is established, considerably less of the blue component will be required in order to establish the match. The amounts of the red and green components necessary for the match will be decreased somewhat, but the ratio of their amounts will be changed only slightly if at all. If the yellow glass is very selective, daylight passed through it may be matched by a simple mixture of the red and green components. In this case no blue component is necessary for the match and the recorded amount of the blue component is zero. If the yellow glass is so selective that the purity of the light passed through it cannot be equalled by any mixture of the red and green components, then it is necessary to establish the match between a mixture of the yellow light with the blue component and a mixture of the red and green components. In this case, the amount of the blue component is considered to be a negative quantity, and is treated accordingly in the computation of the net total of the amounts of the three components, and the coefficients which specify the chromaticity.

In summary, when the purity of the yellow light increased, the coefficient of the blue component necessary for a color match decreased until finally for the purest yellow colors the coefficient of blue component necessary was less than zero, or negative. This negative coefficient is experimentally determined by establishing a color match between a mixture of the yellow with the blue and a mixture of the green with the red components. In a similar manner almost all colors having purities as high or nearly as high as the purity of light from narrow regions of the spectrum require negative coefficients corresponding to one of the components. Negative amounts and negative coefficients of components in color matching always refer to the establishment of a match by mixing the numerically specified amounts of the components with the sample light. In a few rare instances it would conceivably be necessary to establish the color match between one of the components and a mixture of the sample light with the other two components. Such would be the case if the green component were noticeably less pure than a similar green sample, or if the red component were noticeably yellower and less pure than a red sample, or if the blue component were noticeably less pure than a violet sample. The trichromatic components are usually chosen so as to minimize

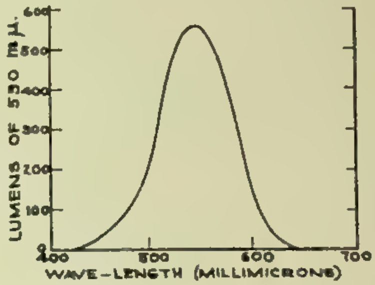
the occurrence of such cases. However, in such cases, the amounts and coefficients of the components which are mixed with the sample are both considered to be negative.

29. THREE-COLOR MIXTURE DATA FOR MATCHING SPECTRUM COLORS

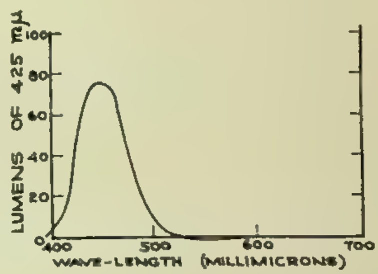
If equal amounts of radiant power from quite narrow regions of the spectrum are employed in turn as samples, the amounts of the red component necessary for the color matches will vary with wave-length in the manner indicated in Fig. 3. This curve indicates that the portions of the curve corresponding to negative amounts join smoothly onto the portions of the curve indicating positive amounts. In a similar manner the curve in Fig. 4 indicates the amounts of the green component necessary for the color matches with equal amounts of radiant power of the several wave-lengths. Finally, the curve in Fig. 5 indicates the amounts of the blue component necessary for the color matches with equal amounts of radiant power of the corresponding wave-lengths. The ordinates at each wave-length of these curves indicate the amounts of the three components in terms of the luminance contributed to the observed mixtures. The standard components for which the color mixture data are shown in Figs. 3–5 consisted of narrow bands of wave-lengths having their average wave-lengths at 650 mμ, 530 mμ, and 425 mμ, respectively. These curves are called color-mixture curves, and refer specifically to the use of the set of trichromatic components described above. Any change in the set of components would result in changes throughout all three of these curves. For any observer, however, the altered curves which result from changes of any or all of the components consist of fractional combinations of three curves such as those shown in Figs. 3–5. For example, the change of the green component from one having a spectral centroid at 530 mμ to one having a spectral centroid at 520 mμ results in a color-mixture curve for the red component consisting of the ordinates of the curve in Fig. 3, plus 4.6 percent of the ordinates of the curve in Fig. 4. The corresponding color-mixture curve for the blue component consists of 0.4 percent of the ordinates of Fig. 4 subtracted from the ordinates of Fig. 5. The altered color-mixture curve for the green component consists of 95.8 percent of the ordinates of the curve shown in Fig. 4. If two of the components are changed, the resulting color-mixture curves for the altered components are fractional combinations of the two corresponding original color-mixture curves. In this case, the third color-mixture curve is a fractional combination of all three original curves. If all three components are changed, then each of the new color-mixture curves is, in general, a fractional combination of all three original color-mixture curves. Such combinations of curves are called linear homo-



3.



4.



5.

FIG. 3. Lumens of 650-mμ component in mixtures with 425- and 530-mμ components matching one watt of each indicated wave-length.

FIG. 4. Lumens of 530-mμ component in mixtures with 425- and 650-mμ components matching one watt of each indicated wave-length.

FIG. 5. Lumens of 425-mμ component in mixtures with 530- and 650-mμ components matching one watt of each indicated wave-length.

geneous combinations, and all color-mixture curves for any specified observer can be computed as linear, homogeneous combinations of any one set of color-mixture curves for that observer.

The example cited above was derived by application of this principle to the data for the standard observer for colorimetry, which will be discussed in Section 34.

30. NON-NEGATIVE COLOR-MIXTURE CURVES

The color-mixture curves corresponding to all possible sets of components indicate negative values for some wave-lengths. However, it is possible to devise linear, homogeneous combinations of such curves which have no negative values. Such combinations result in curves which cannot be obtained in any series of color-mixture observations, yet these curves incorporate and express all of the data represented by the original curves. Such curves are more convenient for computation than are curves indicating some negative values. Since the non-negative curves described are derived from linear, homogeneous combinations of color-mixture curves, it can be said that they are also color-mixture curves, although no possible choice of color-mixture components could result in the direct observation of such curves.

31. ADDITIVITY OF LUMINANCE FOR DIFFERENT COLORS

Within the accuracy of photometric determinations, the luminance produced with a mixture of light from several sources is the sum of the luminances produced by the light from each of the sources acting separately.* Therefore, the sum of the ordinates of the curves shown in Figs. 3–5 is the luminance produced by the fixed amount of radiant power of each wavelength. This is a consequence of the fact that the color matches established in the observations involved photometric luminance matches also. Consequently, the curve representing the sum of the ordinates (expressed in luminance units) of the three color-mixture curves is proportional to the luminosity curve for each observer. This is true of the sums of all sets of color-mixture curves, provided that the ordinates are expressed in terms of equal luminance units.

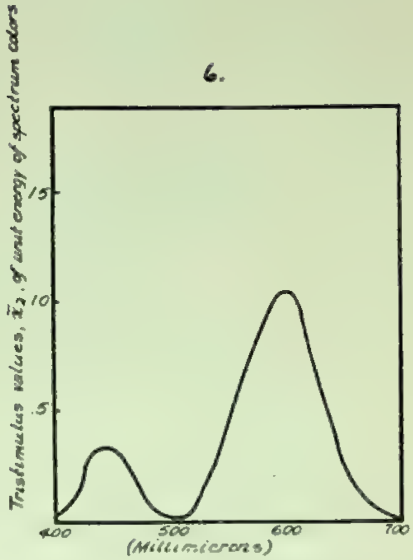
* This statement is true only if each component has sufficient luminance so that photopic vision is employed.

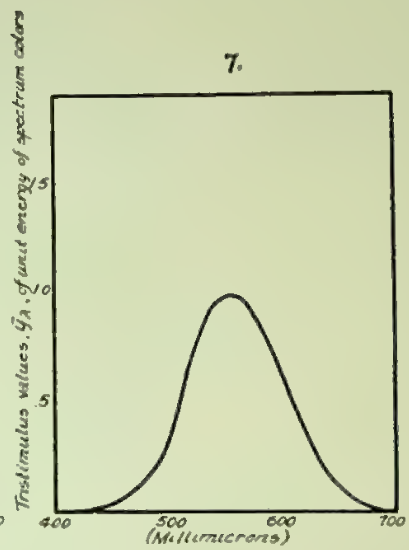
32. LUMINOSITY COEFFICIENTS OF COLOR-MIXTURE CURVES

It is not customary, however, to express the ordinates of color-mixture curves in equal luminance units. Usually these curves are expressed in arbitrary units, different for each of the three curves. When color-mixture curves having no negative values are derived from observed color-mixture curves they are usually expressed in units of such magnitudes that equal areas are enclosed between the base line and the curves, when the data are plotted to equal scales. In this case, the luminosity curve is some linear, homogeneous combination of the three color-mixture curves. The fraction of each color-mixture curve which enters into this linear combination is called the luminosity coefficient of that curve. Actually, the luminosity coefficients are proportional to the luminance of the arbitrary units in terms of which the three-color mixture curves are expressed. Since the luminosity curve is a linear, homogeneous combination of color-mixture curves, it is possible to use the luminosity curve as one of the color-mixture functions in an arbitrarily selected non-negative set. In this case, since one of the color-mixture curves is itself the luminosity curve, the other two color-mixture curves need not and should not contribute in any proportion to the luminosity curve. The corresponding luminosity coefficients would be, therefore, 0, 1, 0. The associated fact that the luminances of the hypothetical red and blue standard stimuli are zero indicates strikingly the unreal character of the stimuli of which the color-mixture proportions are indicated by the red and blue curves. Such hypothetical stimuli may be assumed for the sake of the simplifications thus obtained in the color-mixture curves and their applications.

33. INDIVIDUAL VARIATIONS OF COLOR-MIXTURE CURVES

Since there is such a close connection between the luminosity curve and the color-mixture curves, variations similar in magnitude to the observed variations in luminosity curves must be expected between the color-mixture curves for





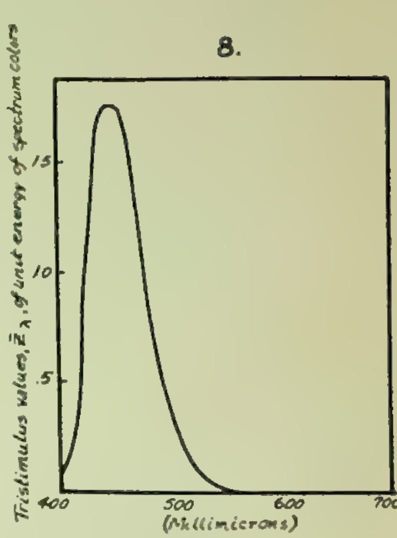


FIG. 6. Standard I.C.I. tristimulus values ($\bar{x}$) of unit energy of indicated wave-lengths.
FIG. 7. Standard I.C.I. tristimulus values ($\bar{y}$) of unit energy of indicated wave-lengths.
FIG. 8. Standard I.C.I. tristimulus values ($\bar{z}$) of unit energy of indicated wave-lengths.

different observers. Furthermore, the effects of selective absorption in the media of the eye and in pigmentation which occurs locally in the retina are variable from one observer to another. Such variations, which may be detected by the well-known "yellow-blue" ratio and "red-green" ratio tests, result in wide variations of the color-mixture curves for different observers.

34. STANDARD COLOR-MIXTURE DATA

The averaged results of two independent investigations of color-mixture curves for a total of seventeen observers[613–615] have been expressed in units consistent with the standard luminosity curve and adopted by the International Commission on Illumination† in 1931.[505] The color-mixture curves which were adopted correspond to hypothetical standard stimuli of the kind described in Section 30, one of the standard color-mixture curves being identical with the standard luminosity curve, the other two color-mixture curves having no negative values. The ordinates of the standard color-mixture curves are expressed in arbitrary units such that the areas under all three curves are equal when the data are plotted to equal scales. The standard curves are exhibited in Figs. 6–8. Their ordinates[505, 616–618] are tabulated in Chapter VII.

Experimental curves consistent with the standard color-mixture curves are found only by rare observers. Such observers can be identified only by selection from a large group of observers, for all of whom color-mixture curves have been determined. Because such a survey of observers is quite impractical, the use of actual observers equivalent to the standard observer for the direct visual determination of standard colorimetric specifications is impractical. The standard color-mixture data can be used for the computation of standard photometric and colorimetric magnitudes. Such computations[617, 618] are based on spectroradiometric and spectrophotometric data. This indirect method of colorimetry is quite practicable, and the computation of standard colorimetric specifications from spectroradiometric and spectrophotometric data is used extensively.

[613] W. D. Wright, Trans. Opt. Soc. **30**, 141 (1928–29).
[614] W. D. Wright, "A re-determination of the trichromatic mixture data," Medical Research Council, Special Report Series, No. 139 (London, 1939).
[615] J. Guild, Phil. Trans. Roy. Soc. **A230**, 149 (1931).
† The abbreviation I.C.I. is customary in America. The initials of the French (CIE) and German (IBK) forms of the name of the commission are customary in European works.

[616] T. Smith and J. Guild, Trans. Opt. Soc. **33**, 74 (1931–32).
[617] D. B. Judd, J. Opt. Soc. Am. **23**, 359 (1933).
[618] A. C. Hardy, *Handbook of Colorimetry* (Technology Press, Cambridge, 1936).

35. SIGNIFICANCE OF STANDARD DATA FOR NON-STANDARD OBSERVERS

Specifications computed by indirect colorimetry permit the classification of ordinary colors in manners corresponding very closely to the appearance of the colors for over ninety percent of the population. The considerable differences which can be detected in the color mixture curves for different observers[619–621] are not sufficient to cause more than ten percent of the population to become aware of any inconsistencies of classifications of ordinary colors based on the standard color-mixture data. About five percent of the population have color vision which is classified as definitely anomalous, since their differences from "normal" observers can easily be detected by the use of simple "color blindness tests."[625–627] The classifications of color based on standard color specifications are usually more complex than anomalous observers consider appropriate. Exceedingly few anomalous observers insist on any essential rearrangement of the standard classifications. Arrangements of colors made by observers with anomalous color vision usually consist of random and unreproducible rearrangements of colors within certain of the groupings defined by the standard data. Such unreproducible rearrangements are evidence that the observers cannot distinguish, and therefore confuse, the colors within these groups which a normal observer can distinguish and arrange.[622, 628, 629]

[619] W. O'D. Pierce, "Individual differences in normal colour vision," Medical Research Council, Special Report Series, No. 181 (1933).

[620] W. D. Wright and F. H. G. Pitt, Proc. Phys. Soc. **47**, 205 (1935).

[621] J. H. Nelson, Proc. Phys. Soc. **49**, 332 (1937).

[622] F. H. G. Pitt, "Characteristics of dichromatic vision," Medical Research Council, Special Report Series, No. 200 (1935).

[623] J. H. Nelson, Proc. Phys. Soc. **50**, 661 (1938).

[624] Winfred M. McKeon and W. D. Wright, Proc. Phys. Soc. **52**, 464 (1940).

[625] S. Ishihara, *Tests for Colour-Blindness* (Kanehara, Tokio, 1930).

[626] E. B. Rabkin, "Polychromatic plates for color sense examination," State Medical Publishing Board (Kief, U.S.S.R., 1936).

[627] American Optical Company, *Pseudo-Isochromatic Plates for Testing Color Perception* (Beck Engraving Company, New York, 1940).

[628] W. O'D. Pierce, *The Selection of Colour Workers* (Pitman, London, 1934).

[629] Institute of Paper Chemistry, Report on Color Blindness (1941).

36. COLOR-MIXTURE DATA ARE INDEPENDENT OF ALL THEORIES OF COLOR VISION

Theories of color vision purport to explain the phenomenon in terms of retinal structure and function, nerve action, and cerebral projection. Color-mixture data on which are based computations of color specifications are independent of all theories of color vision. Discussions of colorimetry are frequently confused by the association of color-mixture facts with the earliest modern theory of color vision, that of Young and Helmholtz. In reality, the laws and data of color mixture are completely independent of any theory, and no reference need be made to any theory of color vision in expositions of the science of colorimetry. On the other hand, no theory of color vision merits consideration which does not fully and accurately account for the facts of color mixture. The Young-Helmholtz theory is commonly linked with the color-mixture data because it accounts for the color-mixture curves in the most straightforward manner. Such simplicity is no guarantee of correctness, however, because other more elaborate theories also consistent with the color-mixture data have been devised. The "Razor of Occam"* which is sometimes employed in favor of the Young-Helmholtz theory of color vision can be invoked further to eliminate *all* theories of color vision from discussions of colorimetry.

37. TRISTIMULUS VALUES

The tristimulus values of any sample of light are defined as the magnitudes of the three standard stimuli computed to be necessary in a mixture matching the light. It is always necessary to specify the units in which these tristimulus values are expressed. This can easily be done by stating in the same units the tristimulus values for some standard or physically defined sample. Such specification is always advisable because the absolute values of spectroradiometric quantities are rarely given or used, and consequently the tristimulus values can only rarely be based on absolute units. For this same reason, the

* *Entia non sunt multiplicanda praeter necessitatem.* This criterion of the medieval controversialist, William of Occam, that in rational discussions no more hypotheses should be employed than are necessary to deal with the facts, has long been a guide for the economy of thought.

conversion to absolute luminosity values, by the use of the factor 650 lumens per watt of radiant power having 555 $m\mu$ wave-length, is rarely encountered in the calculation of tristimulus values.

38. COMPUTATION OF TRISTIMULUS VALUES OF ANY COLOR

Each of the three tristimulus values of a sample of radiant energy may be computed by integrating the product of the ordinates of one of the standard color-mixture curves multiplied by the energy in the corresponding wave-length regions of the spectrum of the sample of radiant energy. If the sample of radiant energy is that reflected from a surface, the spectral composition of the reflected energy is of course given by the products at each wave-length of the spectral

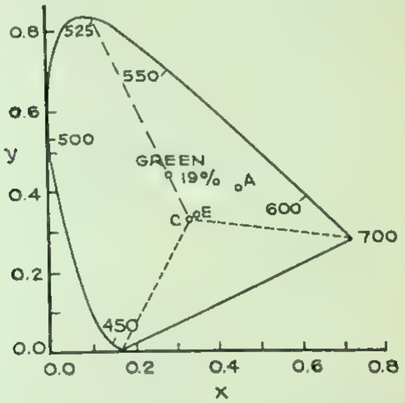


FIG. 9. Chromaticity diagram, illustrating graphical determination of dominant wave-length and purity.

composition of the incident energy and the corresponding directional radiant reflectance of the surface. This product can be employed as the spectral composition of the reflected energy in the calculation of the tristimulus values of the color of the light reflected from the surface. It is customary to express the tristimulus values of reflected light in units such that the tristimulus value resulting from the integration with the luminosity curve is numerically equal to the directional luminous reflectance of the surface for the spectral distribution of radiant energy assumed in the calculations.

The tristimulus values of transmitted light are computed and expressed in similar fashion.

The integration may be carried out by use of the ordinary methods of numerical integration.[617, 618] Special methods[618] have been devised to facilitate the calculations of tristimulus values of the light reflected and transmitted by objects illuminated with light from the most frequently encountered sources. These methods will be explained in detail with examples in Chapter VII of this report, where the necessary accessory data based on the standard color-mixture and energy-distribution data will be tabulated.

39. SPECIFICATION OF CHROMATICITY IN CHROMATICITY DIAGRAM

Chromaticity can be specified by use of the proportions which two of the tristimulus values bear to the total of the three tristimulus values. These two proportions can be employed as coordinates for the representation of the chromaticity of a sample by the position of a point in a plane diagram. Such a diagram is called a chromaticity diagram. In particular, the chromaticity of each spectrum color can be computed by dividing the ordinate of each of the color-mixture curves at the appropriate wave-length by the sum of the ordinates of the three curves at that wave-length. The chromaticities of the various spectrum colors are indicated by the curve in the chromaticity diagram shown in Fig. 9. The horizontal coordinate representing each spectrum color is the ratio of the ordinate at the appropriate wave-length in Fig. 6 to the sum of the ordinates at that wave-length in Figs. 6–8. The vertical coordinate is the ratio of the ordinate in Fig. 7 at the indicated wave-length to the sum of the ordinates. It will be seen that the curve defined by the coordinates representing the chromaticities of energy of the various wave-lengths is smooth and continuous, and that its extremities are quite far apart. The point representing the chromaticity of the light from a source radiating equal amounts in equal intervals of wave-length throughout the spectrum has the coordinates 0.3333, 0.3333. These coordinates result from the fact that the integrals which give the tristimulus values of such light are proportional to the areas under the three color-mixture curves, which three areas are all equal. The chromaticity of such a source is represented by the point labeled *E*. The chromaticity of the standard illuminant which is a satisfactory substitute for average daylight is represented by the

point labeled *C*. The point representing the chromaticity of the light from the I.C.I. standard illuminant *A* (a tungsten incandescent lamp) is labeled *A*. These standard illuminants will be specified completely in Chapter VII of the report.

40. DESCRIPTION OF COLOR MIXTURES IN TERMS OF CHROMATICITY DIAGRAM

The tristimulus values of a mixture of several varieties of light are the sums of the corresponding tristimulus values of the components of the mixture. It follows from this fact that the point representing the chromaticity of a mixture of two components is located on the straight line connecting the points representing the chromaticities of the separate components. The point representing the chromaticity of the mixture is the center of gravity of a system of weights proportional to the sums of the tristimulus values of each component. These weights are located at the points representing the components of the mixture. Consequently, it can be appreciated that the chromaticity of a mixture of daylight with energy taken from a narrow region of the spectrum, at 525 mμ, for example, must be represented by a point somewhere on the straight line between the points representing the chromaticity of daylight (at *C*) and the point representing the chromaticity of the pure spectrum energy (labeled, for this example, with the wave-length 525 mμ). Furthermore, the greater proportion of daylight present in the mixture, the closer the center of gravity (and therefore the point representing the chromaticity of the mixture) will be to the point *C*. Impure colors such as those perceived as pale or pastel are consequently represented by points in the neighborhood of the point representing the chromaticity of daylight. Purer colors are represented by points more remote from *C*, approaching the spectrum locus when the purity of the radiant energy approaches that of the spectrum. Since no light can be purer than spectrum light, there are no samples of light having chromaticities represented by points beyond the spectrum locus. Colors represented by points on the spectrum locus will be called spectrum colors even though, in some cases, the stimulus is not radiant energy confined to one small wave-length interval. The standard stimuli are represented by the points at the intersection of the coordinate axes and at the coordinates (0, 1) and (1, 0). These points lie outside the area enclosed by the spectrum locus and the straight line connecting its extremities. There are no samples of light having the chromaticities of these standard stimuli. Mixtures having the chromaticities represented at any points along the straight line joining the extremities of the spectrum locus can be produced by mixing suitable proportions of radiant energy taken from the extreme short wave (less than 440 mμ) and long wave (greater than 680 mμ) regions of the visible spectrum. Such mixtures are non-spectral colors and can be arranged in order of hue according to the ratio of intensities of the long wave and short wave components. The purest non-spectral colors are represented by points on this straight line connecting the extremities of the spectrum locus. Less pure non-spectral colors are represented by points lying closer to the achromatic point *C*. Colors represented by points on straight lines between the achromatic point and the spectrum locus will be called spectral colors. The spectrum locus, together with the straight line joining its extremities, may be called the boundary of real colors.

41. ACHROMATIC STIMULI

An alternative numerical system for specifying chromaticity is based on the position of the point representing chromaticity in the diagram. The use of this system of specification requires a decision as to what kind of light will be said to have zero purity, that is, what kind of light shall be considered achromatic. Since the specification of chromaticity which is based on this decision has reasonableness of correlation with ordinary color description as its chief merit, the choice of the achromatic stimulus should always be such as to yield the most reasonable specifications. In most ordinary circumstances, the chromaticity of the prevailing illuminant is a satisfactory choice for the achromatic stimulus. For brevity, the point representing the achromatic stimulus may be called the achromatic point. Similarly, the point representing the sample may be called the sample point.

42. DOMINANT WAVE-LENGTH AND PURITY

When the achromatic stimulus has been decided upon, the chromaticity can be specified in terms of dominant (or complementary) wave-length and purity. The *dominant* wave-length of a spectral color is the wave-length associated with the point on the spectrum locus which also lies on the extension of the straight line drawn through the sample point from the achromatic point. In the case of non-spectral colors, the line drawn in this manner does not intersect the spectrum locus. In this case the chromaticity is specified by *complementary* wave-length and purity. The complementary wave-length is defined as the wave-length associated with the point on the spectrum locus which also lies on the straight line drawn from the sample point and extended through the achromatic point. The purity of any sample of light is determined by the distance of the sample point from the achromatic point. The purity is equal to this distance, expressed as a fraction of the distance between the achromatic point and the intersection of the boundary of real colors with the line drawn from the achromatic point and extended through the sample point. In Fig. 9 the point labeled "green" corresponds to a purity of 19 percent. This definition of purity conforms with the definition of the quantity frequently called excitation purity, as distinguished from colorimetric purity. Since the use of a single word is desirable wherever possible, the unmodified word purity will refer to the concept defined above, otherwise known as excitation purity. This choice between the two measures available for purity is based primarily on considerations of computational convenience, each of the measures of purity being arbitrary, and neither being related in any simple way to visual sensitivities to purity change. Formulas and tables will be given in Chapter VII for the conversion of colorimetric purity to excitation purity, and the reverse.

It is permissible and sometimes desirable to determine purities based on chromatic rather than achromatic reference stimuli. Anomalous expressions can be avoided in such cases by substituting the terms *reference stimulus* and *reference point* in place of the more specific terms, achromatic stimulus and achromatic point.

43. INTERPRETATION OF CHROMATICITY DIFFERENCES

Although specifications of colors are easily interpreted in terms of dominant wave-length and purity, a supplementary specification of the kind and degree of chromaticity difference is sometimes convenient. The kind of chromaticity difference between a sample and a reference stimulus may be specified by the wave-length (or complementary wave-length) corresponding to the point at which the straight line drawn from the reference point and extended through the sample point intersects the boundary of real colors. This wave-length has been named *conjunctive wave-length*.[630] The degree of color difference may be specified by the purity of the sample

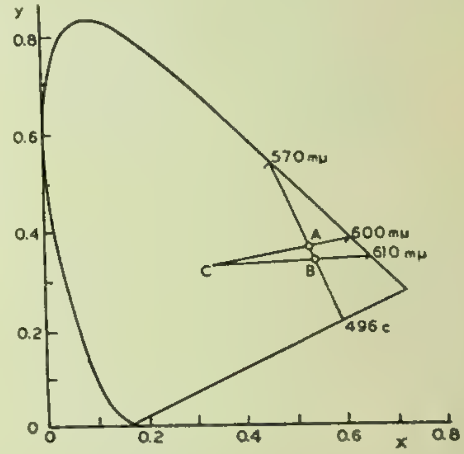


FIG. 10. Chromaticity diagram, illustrating graphical determination of conjunctive wave-length for specification of color difference.

computed with respect to the reference stimulus. This specification of a chromaticity difference supplements, but does not replace the specifications of the two colors in terms of dominant wave-length, purity, and luminance.

An example of this supplementary specification of chromaticity difference is illustrated in Fig. 10. Sample A has a dominant wave-length of 600 mμ and excitation purity 70 percent. Sample B has dominant wave-length 610 mμ, purity 65 percent, and the same luminance as A. The straight line joining A and B intersects the spectrum locus at about 570 mμ and the boundary of non-spectral colors at about 496 c. The first is

[630] R. Davis, Bur. Stand. J. Research 7, 659 (1931).

the conjunctive wave-length of A with respect to B. The second is the conjunctive complementary wave-length of B with respect to A. The purity of A with respect to B (in this case about 12 percent) is, within limits, proportional to the noticeability of the specified kind (570 mμ) of color difference from B as a reference stimulus. Visual sensitivities to chromaticity differences vary radically with changes of the reference stimulus and conjunctive wave-length.[631, 632]

[631] W. D. Wright, Proc. Phys. Soc. **53**, 93 (1941).
[632] D. L. MacAdam, J. Opt. Soc. Am. **32**, 247 (1942).

Scalar Polarization Fringes Produced by the Superposition of Crystalline Plates

Bruce H. Billings
Research Laboratory, Polaroid Corporation, Cambridge, Massachusetts
(Received February 11, 1944)

A. H. PFUND[1] has described the pattern produced by the superposition of the two sets of Haidinger fringes obtained with a half-silvered mica plate. The locus of the spots where the fringes cross is a new set of fringes which have been called moiré fringes. The name "scalar interference fringes" has been proposed by E. H. Land for the moiré fringes which occur when any two sets of fringes cross.

The scalar fringes which we shall demonstrate were obtained by superposing different sets of polarization fringes in crystalline plates. The arrangement of our system for observing these fringes can be seen in Fig. 1. The circular polarizers shown in this figure are made with Polaroid sheet polarizing material and quarter-wave plates. The three polarizers in Fig. 1 are all of the same sign. That is, if the crystal plates are removed, the transmission is essentially the same as the transmission through the first circular polarizer. Without the central circular polarizer, the two primary sets of fringes are replaced by a series of spots. The location of these spots in the treatment of the general pattern of superposed crystalline plates has been carried through by Pockels[2] and Mascart[3] and many other writers. Excellent pictures can be seen in the compendium by Hauswaldt.[4]

The simplest pattern is probably that produced by the intersection of two sets of circles. Circular polarization fringes can be easily produced with a basal section of a uniaxial crystal between

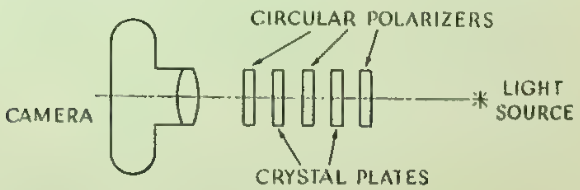


Fig. 1. Sketch of apparatus.

[1] A. H. Pfund, J. Opt. Soc. Am. 32, 383 (1942).
[2] F. Pockels, Göttingen Nachrichten (1890).
[3] M. E. Mascart, *Traité d'Optique* (Gauthier-Villars, Paris, 1891), Vol. II, Chap. X.
[4] Hans Hauswaldt, *Interferenzerscheinungen im polarisirten Licht* (Joh. Gottlieb Hauswaldt, Magdeburg, 1904).

circular polarizers. Figure 2 shows such a set of rings which can be described by the following formula:

$$\sin^2 i = \frac{(2n+1)\lambda}{d}\,\frac{\mu_e^2\mu_0}{\mu_0^2-\mu_e^2}, \tag{1}$$

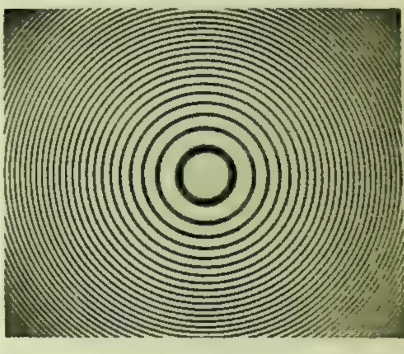


Fig. 2. Single crystal between crossed circular polarizers.

where i is the angle of incidence of light which produces a ring; d is the thickness of the plate; n is an integer; μ_e is the index of refraction for the extraordinary ray; and μ_0 is the index for the ordinary ray. This formula is valid only for small angles of incidence. It is derived by taking the expression

$$\Gamma = \frac{d}{\lambda}(\mu_2 \cos r_2 - \mu_1 \cos r_1) \tag{2}$$

for the relative retardation of the two beams formed when light hits the plate at an angle of incidence i and replacing it with the following formula:

$$\Gamma = \frac{d}{\lambda}(\mu_2 - \mu_1). \tag{3}$$

μ_1 and μ_2 are the indices at angles r_1 and r_2 with the normal.

When two sets of circular rings are superposed, several new sets of fringes are produced. The larger rings in a given set will be weaker than the smaller ones. Therefore, the crossover points of rings of equal order will tend to be equal in intensities. Also, crossover points of rings, the

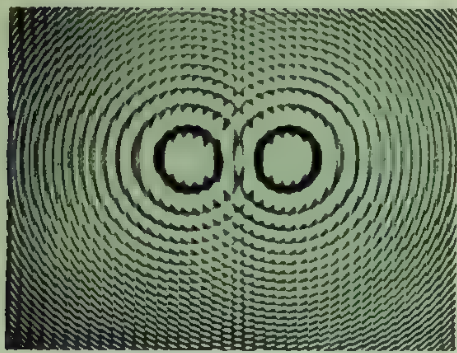


FIG. 3. Fringes produced with two superposed basal sections of a uniaxial crystal. Central ring diameters are 2°, angle between optic axes 3° 15′, and light source sodium arc.

sum of whose orders is constant, should tend to be of equal intensities.

Suppose the centers of our two ring systems are displaced by an angle $2a$, and suppose the origin of our coordinate systems to be halfway between the two centers. We can write the formula of our two series as

$$(x-a)^2+y^2=n_1r_1^2, \tag{4}$$

$$(x+a)^2+y^2=n_2r_2^2, \tag{5}$$

where $r_1^2=\sin^2 i$ and is the square of the radius of the first-order ring in plate number 1; n_1 is an odd number. We can represent all possible sets of scalar interference fringes by eliminating n_1 and n_2 from these equations with the help of the equation

$$\alpha_1n_1+\alpha_2n_2=k. \tag{6}$$

By a suitable choice of α_1, α_2, and k, we can get the formulae for all the different possible scalar interference fringes. By carrying through the elimination, we get

$$x^2-2ax\frac{\alpha_2r_1^2-\alpha_1r_2^2}{\alpha_2r_1^2+\alpha_1r_2^2}+a^2+y^2=\frac{kr_2^2r_1^2}{\alpha_2r_1^2+\alpha_1r_2^2}. \tag{7}$$

We will consider those sets of scalar interference fringes which we have actually observed. Consider first that $r_1=r_2$. For this we have the general equation

$$x^2-2ax\frac{\alpha_2-\alpha_1}{\alpha_2+\alpha_1}+a^2+y^2=\frac{kr_1^2}{\alpha_2+\alpha_1}. \tag{8}$$

If now we let $\alpha_1=-\alpha_2=1$, we get

$$x=\frac{kr_1^2}{4a}. \tag{9}$$

k is now the difference between odd numbers and is, therefore, an even integer. This represents a series of straight lines of separation $r_1^2/2a$.

We can describe still another pattern with the values

$$\alpha_1=\tfrac{1}{2},\qquad \alpha_1=-1,$$

$$\alpha_2=-1,\quad \alpha_2=\tfrac{1}{2}.$$

This gives us the sets of circles

$$(x-3a)^2+y^2=-2kr_1^2+8a^2, \tag{10}$$

$$(x+3a)^2+y^2=-2kr_1^2+8a^2. \tag{11}$$

In Fig. 3 we can see the patterns produced by superposing two sets of equal rings. As predicted by Eq. (9), the picture is crossed with a series of straight lines. The radii of the central rings are 1°. The separation of their centers is 3° 15′. If we put these values in Eq. (9), we get 18′ for the separation of the lines. This agrees with the observed value as closely as it can be measured. At the edges of the picture we can see the faint sets of rings described by Eqs. (10) and (11). The centers of these sets of rings are 4° 52′ from the center of our coordinate system. This distance is $\frac{3}{2}$ the separation between the centers of our polarization rings, as expected from Eqs. (10) and (11).

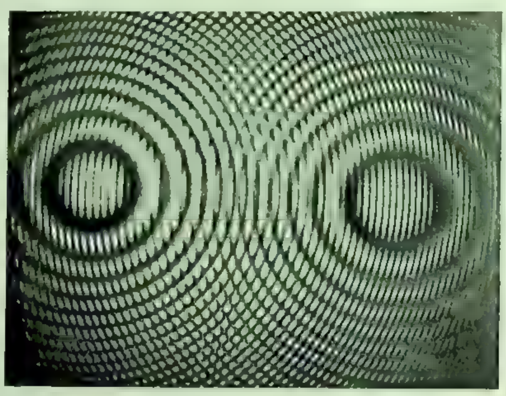


FIG. 4. Fringes produced with two superposed basal sections of a uniaxial crystal. Central ring diameters are 2° 36′ and 3°, angle between optic axes 9°, light source sodium arc.

Consider now the effect in Eq. (8) of letting $\alpha_2=\alpha_1=1$. We now get

$$x^2+y^2=\frac{kr_1^2}{2}-a^2. \tag{12}$$

This represents a set of circles with centers at the origin. These circles can be seen in Fig. 4. The lines which cross them are similar to those in Fig. 3, except that they are slightly curved as the two primary rings are of unequal size.

If the primary rings are quite different in size, the set of straight lines described by Eq. (9) become circles. By putting $\alpha_1 = -\alpha_2 = 1$ in Eq. (7), we get the equation of these circles.

$$\left(x + a\cdot\frac{r_2^2 + r_1^2}{r_2^2 - r_1^2}\right)^2 + y^2 = \frac{kr_2^2 r_1^2}{r_2^2 - r_1^2} + a^2\frac{4r_1^2 r_2^2}{(r_2^2 - r_1^2)^2}. \quad (13)$$

Figure 5 shows this set of circles which was obtained with two primary rings 1° 40′ and 2° 30′ in diameter.

Since these scalar interference fringes are obtained simply by adding our primary sets of fringes, we can use as primary fringes a set similar to one of our scalar patterns and get as scalar pattern a set similar to the original primaries. To illustrate this, we use the fringes in Fig. 4. Here our scalar fringes consist of a set of circles crossed by a set of lines. The primary fringes are two sets of circles. With a single set of circles and straight lines as primary fringes, we get as scalar fringes the original two sets of circles. The lines were produced by two crystal plates between crossed polarizers. The principal planes of the plates were at right angles to each other, and their optic axes made an angle θ with the normal. The equation of these fringes may be written as

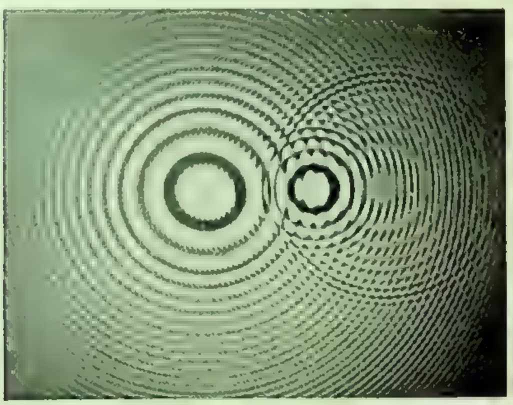


FIG. 5. Fringes produced with two superposed basal sections of a uniaxial crystal. Central ring diameters are 1° 40′ and 2° 30′, angle between optic axes 4° 10′, light source sodium arc.

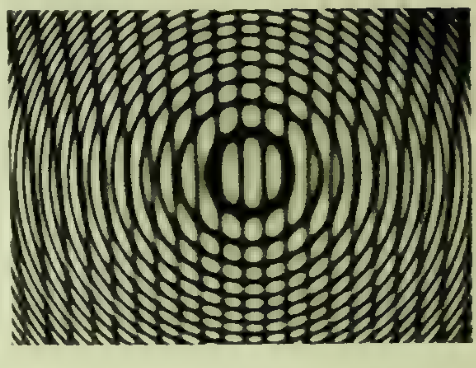


FIG. 6. Fringes produced with superposed basal section of a uniaxial crystal and a Savart plate. Ring diameter 3°, angular separation between Savart plate lines 50′, light source sodium arc.

$$\frac{\mu_0^2 n\lambda}{d(\mu_0 - \mu_e)} = \sin^2\theta \sin^2 i(\sin^2\zeta - \cos^2\zeta) + \mu_0 \sin 2\theta \sin i(\sin\zeta - \cos\zeta). \quad (14)$$

Here ζ is the angle the plane of incidence makes with the principal plane of the first plate. This equation is of a set of hyperbolas. Since we are considering only small angles, we can neglect the first term and we get

$$\frac{\mu_0 n\lambda}{d(\mu_0 - \mu_e)} = \sin 2\theta \sin i(\sin\zeta - \cos\zeta). \quad (15)$$

This represents a set of straight lines. We can replace $\sin i \sin\zeta$ with y, and $\sin i \cos\zeta$ with x. Our equation now reads

$$\frac{\mu_0 n\lambda}{d(\mu_0 - \mu_e)} = \sin 2\theta(y - x). \quad (16)$$

By a rotation of axes, we can put this in the form of Eq. (9), and with Eq. (12) we get Eqs. (3) and (4) for our scalar pattern. In Fig. 6 are shown the fringes obtained with such a system. The straight lines are slightly curved as might be expected from the approximation.

These are, of course, only some of the simplest of the patterns that can be produced. It is hoped that at some later date a more thorough investigation can be carried through.

New Catadioptric Meniscus Systems

D. D. MAKSUTOV
State Optical Institute, Moscow, U.S.S.R.
(Received January 26, 1944)

1. PREFACE

THE present paper contains an abridged account of the author's work under the same heading on new optical systems of a type invented by him in August, 1941.*

The defects of modern refractors and reflectors induced the author to look for a new solution. Such a solution has been found in the form of meniscus catadioptric systems. These systems are practically free from chromatism, spherical aberration, coma, and in many cases also from astigmatism. The instruments of the new type possess a high degree of compactness, are simple and cheap to construct, have a large relative aperture, high optical quality, and finally are reliable in operation. Any of the classical systems of telescopes may be transformed into a meniscus one with great advantages at the same time.

The high universality of the new systems assures their successful use in almost every branch of optical instrument making (geodesic, photographic, spectral, laboratory instruments, microscopes, and so on). However, this paper is written with the interests of astronomers in mind and will afford them the possibility of designing and making new instruments for amateur observations as well as for scientific ones. The present paper gives only the principal results, without theoretical derivations and details of design.

Subjecting his invention to broad publicity, the author is glad to share it with American and English colleagues, and hopes to stimulate in that way the general progress of astronomical science and to strengthen the cultural relations between our countries.

2. THE DEFECTS OF REFRACTORS AND REFLECTORS

The Refractor

To begin with, the refractor is not sufficiently achromatic. To fulfill in an achromatic objective the condition of Rayleigh (a quarter wave-length for wave aberrations) for rays C and F, the following relation has to be satisfied:

$$D_{\max}=\frac{2.22}{A}=2.22V, \tag{1}$$

where $D_{\max}$ is the maximum diameter of the achromat, expressed in millimeters, $A=D:f$ is the relative aperture of the objective, and $V=f:D$ is the inverse quantity or the relative focal length.*

Thus an achromat with diameter $D=100$ mm must have $A=1:45$ and $f=4.5$ mm; an objective (of the Yerkes type) with $D=1000$ mm must have $A=1:450$ and $f=0.45$ km! Owing to economic and technical considerations the construction of instruments of such a length is not possible; hence refractors, not only of great but of medium diameter as well, have an excessive secondary spectrum that obviously makes the image quality worse, this effect being the greater the wider the spectral range used. We are *nolens volens* compelled to tolerate this defect.

Half-achromats and apochromats also do not give a practically useful solution of the problem. In the first place, only a relatively small decrease of the secondary spectrum is obtained; secondly, the optical constants of glass have values requiring considerable lens curvatures, the result being an excessive increase of spherochromatic aberration. Thus, an apochromat with an increased relative aperture turns out to be a system chromatically less perfect than an achromat of the same relative aperture, notwithstanding the apparent advantage of the decrease of the secondary spectrum.

Refractors, then, inevitably must be either very long and cumbersome instruments or optically imperfect systems.

A refractor with an aperture of 400 mm is probably the limit from the point of view of reasonable expense and not too much trouble; at the same

* Patented in U.S.S.R. according to the claim dated November 3, 1941.

* The last two notations, characteristic of the relative aperture of the optical system, will be encountered often in the following exposition.

time its relative aperture must be diminished to at least 1 : 40 ($f \sim 16$ m) instead of the usually admitted value 1 : 15–1 : 20, in order to obtain a tolerable residual chromatism.

A direct consequence of chromatism is the impossibility of constructing the powerful objectives having at the same time the high relative aperture that is required by stellar photography.

Other defects of objectives, connected with the inhomogeneity of glass, with the light absorption in the short region of the spectrum, with the technology of construction, etc., will not be considered here.

The Reflector

Each reflector requires the use of aspherical surfaces of high precision, difficult both to make and to investigate. If the increased precision of making reflecting surfaces is taken into account, it is easy to understand that the construction of first-rate reflectors with high relative aperture often lies outside the limits of capability of the optician and the possibilities of methods of control. In addition, reflectors of the classical systems have a non-corrected coma, quite inadmissible for astronomical instruments used to resolve different problems.

Aplanatic telescopes of Schwarzschild, Chrétien, the author,[1] and Coudé afford the possibility of correcting coma. However, with the exception of the author's systems with concave elliptic mirrors, they require a form of mirror for which it is not possible to find a simple, convenient, and sensitive method of control. The open tubes of reflectors are favorable for convection streams of air along the path of the rays and for stronger temperature influences on the mirrors; owing to this the image quality becomes considerably worse. Moreover, due to the open tubes, the reflecting metallic layers become dusty and dim and are spoiled.

The necessity of mounting the secondary mirrors on supports across the path of the rays leads to harmful diffraction effects. The compromise solution suggested by the author[2] in the form of bent and beaded supports affords the elimination of diffraction rays for star images only by their distribution over 360° instead of several directions; i.e., it means only the change in the external form of diffraction hindrances, but not in their essential properties. Besides this the fastening of secondary mirrors in settings and on supports has as a result an additional screening and tends to throw the instrument out of adjustment.

Other defects of reflectors will not be considered here.

Among mixed or catadioptric systems only two are worthy of serious consideration, that of Schmidt and that of Ross. Each of them requires the use of aspherical optics, i.e., has the chief defects of the reflector, and apart from this has some specific defects of less importance that will not be analyzed here.

Thus the designer of astronomical instruments is in a quandary: Either he chooses the refractor and tolerates a cumbersome and expensive instrument[3] having comparatively small power and relative aperture and at the same time considerable chromatic hindrances, or he selects the reflector with difficultly realizable aspherical surfaces of high precision but inherent defects. The best illustration of the above considerations is afforded by the history of competition between the refractor and the reflector and of the struggle of their advocates, that has not yet ceased.

3. THE PRINCIPLE OF MENISCUS SYSTEMS

Meniscus systems make possible the elimination of the chief defects of the reflector and the refractor and a considerable decrease of their defects of minor importance. The principle of meniscus systems may be stated as follows.

A concave spherical mirror or a telescopic system consisting of centered spherical mirrors is perfectly achromatic, but has a considerable negative spherical aberration, and coma. The so-called achromatic meniscus, i.e., a lens from optical glass of a nearly constant thickness, but with considerable curvatures of surfaces suitably chosen, is practically free from chromatism, but has a considerably positive spherical aberration. By combining an achromatic meniscus with a

[1] D. Maksutov, Transactions of the State Optical Institute, Part 86, 1932.

[2] D. Maksutov, Proc. Acad. Sci. U.S.S.R., p. 521 (1937). Patent N 49359, dated September 29, 1936.

[3] The cost of the building and of the dome increases proportionally to at least the square of the instrument's length.

spherical mirror it is possible to correct the spherical aberration without bringing in a noticeable chromatism.

Changing the distance between the meniscus and the principal mirror, or changing some of the parameters in more complex systems, it is possible in addition to eliminate the coma and often the astigmatism also, the last correction being unnecessary in most instruments.

In the case of a two-lens objective the spherical aberration of the positive lens is compensated by the spherical aberration of the inverse sign of the negative lens; in the case of a meniscus system a compensation principle can also be applied for the correction of spherical aberration. But in nearly all compensation schemes it is possible to correct the aberration only in the first approximation; hence in meniscus systems as in objectives, a residual aberration remains which determines the limits of increase of the diameter or of the relative aperture of the system. Fortunately these limits for meniscus systems prove to be very favorable.

The same must be said about the residual chromatism, which is present in meniscus systems, but is a hundred times smaller than that for equivalent objectives.

A meniscus, achromatic for paraxial rays, has to satisfy the following condition:

$$\Delta R/d=(n^2-1)/n^2, \tag{2}$$

where $\Delta R=R_1-R_2$ is the difference between the radii of the first and the second meniscus surfaces and n is the refractive index of the meniscus substance.

The inverse focus of the meniscus will be

$$\varphi_0=\frac{I}{f_0}\cong-\frac{d}{R_1^2}\,\frac{(n-1)^2}{n^2} \tag{3}$$

and the angular spherical aberration for the zone y due to the meniscus will be

$$\eta_y\cong+\frac{y^2d}{2R_1^4}\,\frac{(n^2-1)(n-1)}{n^2}. \tag{4}$$

Thus the achromatic meniscus is a negative lens ($\varphi_0<0$) of a very long focus, giving a considerable spherical aberration ($\eta_y>0$).

The achromatic meniscus is not free from chromatism, but owing to a very small relative aperture, φ, the secondary spectrum of the meniscus system is 500–800 times smaller than the secondary spectrum of an equivalent achromatic objective.

If we take for a system with a finite aperture a meniscus achromatic for paraxial rays, a noticeable chromatism in the external zone appears, due to spherochromatic aberration. Therefore, in a meniscus of a real system some change of condition (2) in the direction of greater values has to be made.

As is seen, an achromatic meniscus may be made from glass with any n value. Of course, it is advantageous to use, for a meniscus, glass with small dispersion (crowns), since the residual chromatism depends on the dispersion; though in meniscus systems this is generally small, there is no reason to increase it unnecessarily.

In particular, the meniscus may be made of uviol crown or of fused quartz, which makes it possible, in combination with aluminized mirrors, to construct an instrument for the ultraviolet region of the spectrum.

The principal role of the meniscus is the correction of spherical aberration and coma of spherical mirrors, without introducing a noticeable chromatism; however, in addition to that, the meniscus fulfills two very important functions. Firstly, it shuts the tube and makes possible a sealed instrument, free from convection streams, with mirrors well protected from dust, damage, and the influence of abrupt changes of temperature; secondly, the meniscus makes possible the fastening to it of secondary mirrors without supports, and often also without settings by means of an adhesive or by optical contact; in the most special case the central part of the inner surface of the meniscus may be aluminized and plays the role of a secondary mirror, the screening and the diffraction hindrances being in that case very small.

Meniscus systems have the following advantages:

(1) They assure a high image quality for a wide wave-length range.
(2) For ultraviolet rays they are much more transparent than the objectives, since they do not require flint lenses.
(3) Being a combination of spherical optics they are very simple and cheap to make and their testing is simple and precise.

(4) They make possible the construction of instruments having high relative apertures and wide fields and being at the same time compact, light, and easy to manipulate.
(5) They form a sealed system with all its advantages and with small possibilities of disadjustment.
(6) The advantageously resolved problem of optical glass affords its choice not only according to optical, but also to physico-chemical, technological, and considerations of design.

The idea of meniscus systems is so simple that it is surprising that during more than three centuries of optical instrument making it did not occur to anyone, though in its essential part it was available to the contemporaries of Descartes.

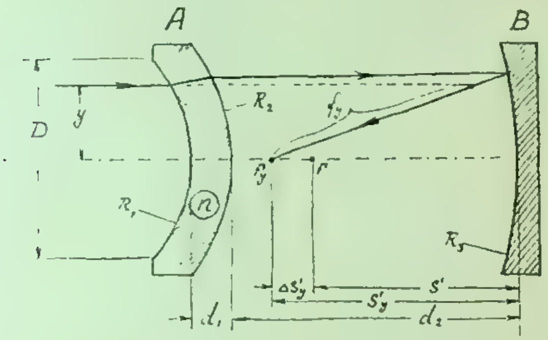


FIG. 1.

4. "MENISCUS-CONCAVE MIRROR" SYSTEM

It is most appropriate to consider the properties of meniscus systems in general by using the example of the simplest and at the same time most fundamental system.

Figure 1 shows such a system, consisting of a meniscus A and a lens B. The meniscus system may be determined by the diameter of the aperture D, the radii of curvature of the meniscus and of the mirrors R_1, R_2, R_3 (in the case of Fig. 1 all three radii are negative), the meniscus thickness d_1, the refractive index of its material **n**, and the distance d_2 between the meniscus and the mirror.

All this being given, the last interval S_y' and the focal length f_y fór each zone **y** are determined in a unique way. If the focus of the zone **y** is in the point F_y and the focal plane goes through the point F_y, the logitudinal aberration $\Delta S_y'$ with respect to the focal plane determines the angular aberration $\Delta\eta_y$ for the same zone **y** in the following way:

$$\Delta\eta_y = \frac{y \cdot \Delta S_y'}{f^2}. \tag{5}$$

At $\Delta S'_{y_i}=0$ or at $\Delta\eta_{y_i}=0$, the system becomes stigmatic for the focus F. If at the same time $f_{y_i}=\text{const.}=f$, the system becomes aplanatic, i.e., simultaneously corrected for spherical aberration and for coma. It is the solution aimed at; if this aim is not attained completely we have a system with residual spherical aberration and residual coma. The fulfillment of the condition of isoplanatism affords a most advantageous solution.

If for different values of the refractive index n_λ, corresponding to different wave-lengths λ, $S_\lambda'=\text{const.}$ ($\Delta S_\lambda'=0$ or $\Delta\eta_\lambda=0$), the system is free from chromatism of position; if that condition is not satisfied for all zones there is a spherochromatic aberration in the meniscus system.

If $f_\lambda=\text{const.}$, the system is free from chromatism of magnification; otherwise there is a chromatic difference of magnification in the system.

The problem for the calculator is reduced to finding a system as achromatic and aplanatic as possible, i.e., to reducing residual aberration to a minimum.

A great number of trigonometrical computations made by the author afford empirical computation formulae for one special case of the meniscus-concave mirror system. This special case is limited by three conditions:

(1) For the meniscus the glass is taken: $n_D=1.5163$; $\nu=64.1$.
(2) The relative thickness of the meniscus is $d_1 : D=0.1$.
(3) The condition of the minimal circle of diffusion is satisfied.

The first two of these restrictions suggest to a certain degree the kind of glass and the relative meniscus thickness.

The empirical formulae afford a rapid solution near the optimal one; a subsequent trigonometrical computation either confirms the correctness of the solution or gives some insignificant corrections.

TABLE I. [$S'=419.983$].

Zone y	0		17.5		30		42.5		50	
Ray	C	F	C	F	C	F	C	F	C	F
S_y'	420.000	420.008	419.991	419.998	419.981	419.985	419.978	419.977	419.989	419.985
$\Delta S_y'$	+0.017	+0.025	+0.008	+0.015	−0.002	+0.002	−0.005	−0.006	+0.006	+0.002
$\Delta\eta_y$	±0″	±0″	+0.18″	+0.34″	−0.08″	+0.08″	−0.27″	−0.33″	+0.39″	+0.13″
$\Delta\eta_{C,F}$	±0″		−0.16″		−0.16″		+0.06″		+0.26″	
f_y	400.109	400.018	400.121	400.029	400.138	400.049	400.157	400.067	400.156	400.070
Δf_y	±0	−0.091	+0.012	−0.080	+0.029	−0.060	+0.048	−0.042	+0.047	−0.039

The following empirical formulae allow the determination of the all main optical elements of the system with an aperture D, i.e., the determination of the meniscus radii R_1 and R_2, the mirror radius R_3, and the distance d_2 between the meniscus and the mirror:

$$R_1 : D = -0.612V^{0.66}, \tag{6}$$

$$R_2 : D = -0.612V^{0.66} - 0.0565 - 0.007A, \tag{7}$$

$$R_3 : D = -2.107V^{0.983}, \tag{8}$$

$$d_2 : D = 1.11V^{1.14}, \tag{9}$$

the relative thickness of the meniscus taken equal to

$$d_1 : D = 0.1 \tag{10}$$

and the kind of glass being specified by the constants

$$n_D = 1.5163; \quad \nu = 64.1.$$

Let us apply these formulae to the calculation of a system, for which, e.g.,

$$D = 100 \text{ mm}; \quad A = 1:4.$$

The computation with the help of formulae (6)-(10) leads to the following value of the optical elements:

$$\begin{aligned} D&=100, & R_1&=-152.8, & d_1&=10.0,\\ A&=1:4, & R_2&=-158.6, & d_2&=539.1.\\ && R_3&=-823.2, && \end{aligned} \tag{11}$$

The trigonometrical computation of such a system gives the following distances and residual aberrations for five zones **y** and for two colors C and F. See Table I.

It is seen that for the most advantageous focusing ($S'=419.983$), the longitudinal aberrations $\Delta S_y'$ in the outer zones of the system do not exceed several microns. Correspondingly, the angular aberrations for neither zone exceed 0.4″ (arc seconds) for the spectral range from C to F; for monochromatic radiation they are still less. The residual chromatism $\Delta\eta_{C,F}$ due to the spherochromatic aberration lies within the limits −0.16″, +0.26″.

The focal lengths f_y for different zones in monochromatic light are nearly constant; this indicates a very good approximation to complete aplanatism.

For rays C the focal lengths are 0.09 mm longer than the focal lengths for rays F. Hence, with $f=400$, the chromatic difference of magnification

$$\frac{f_C - f_F}{f} \cdot 100 \text{ percent} = 0.023 \text{ percent}, \tag{12}$$

i.e., in that respect also the system is quite satisfactory.

A system computed with the help of empirical formulae has proved to be so near to the optimum that it requires almost no further improvement.

Figure 2 shows the curves $\Delta S_y'$ and $\Delta\eta_y$. The curve $\Delta\eta_y$ goes through a maximum ($\Delta\eta_{max}$) in the zone $y \sim 0.35(D/2)$, then through a minimum in the zone $y \sim 0.85(D/2)$ and attains again a maximum in the outer zone $D/2$. If, for a certain ray λ, both maxima are numerically equal to the minimum, then for that ray the condition of minimal circle of diffusion is satisfied; the same condition turns out to be sufficiently near, although not identical with the condition of minimal wave aberrations. Figure 3 shows the curves Δf_y of the system for rays C and F and also the dotted curve Δf_y for the equivalent parabolic mirror. The comparison of these curves shows the smallness of the residual coma of the meniscus system, compared with the coma of a parabolic mirror of the same diameter and of the same relative aperture as the meniscus system considered.

This example shows the possibility of realizing meniscus systems that are in a high degree achromatic, aplanatic, and that have a sufficient relative aperture.

Let us now consider what limits are reached for the relative apertures and diameters of meniscus systems without going outside the Rayleigh condition of one-quarter wave-lengths for wave aberration in monochromatic light ($\lambda \sim 555$ mμ).

5. THE LIMITING DIAMETERS AND RELATIVE APERTURES

A considerable number of trigonometrical computations have allowed us to find the laws of variation of residual aberrations in dependence on the variation of the relative aperture A in meniscus-concave mirror systems, calculated with the help of formulae (6)–(10), and to obtain empirical formulae for an approximate computation of residual aberrations.

Thus the maximal residual aberration $\Delta\eta_{\max}$ is approximately proportional to the degree 4.5 of the relative aperture of the system and may be

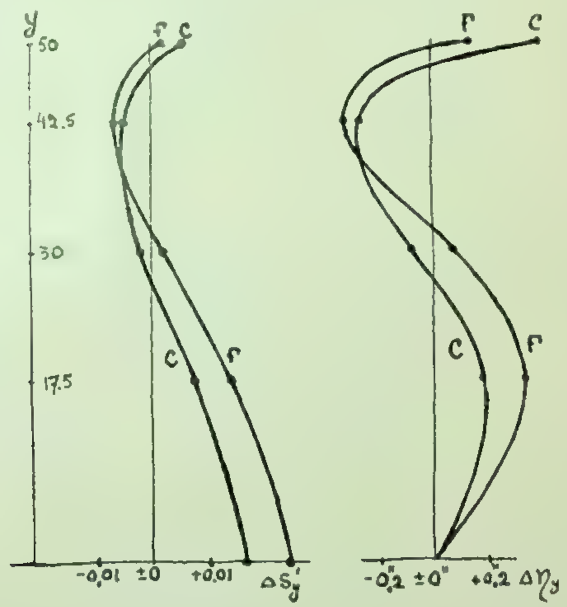


FIG. 2.

expressed by the formula

$$\Delta\eta_{\max} = 123A^{4.5} \text{ (arc seconds)}. \qquad (13)$$

The residual chromatism (the spherochromatic aberration) $\pm\Delta\eta_{C,F}$ is approximately proportional to the cube of the relative aperture and may be expressed by the formula

$$\pm\Delta\eta_{C,F} = 7.6A^3 \text{ (arc seconds)}. \qquad (14)$$

The chromatic difference of magnification may

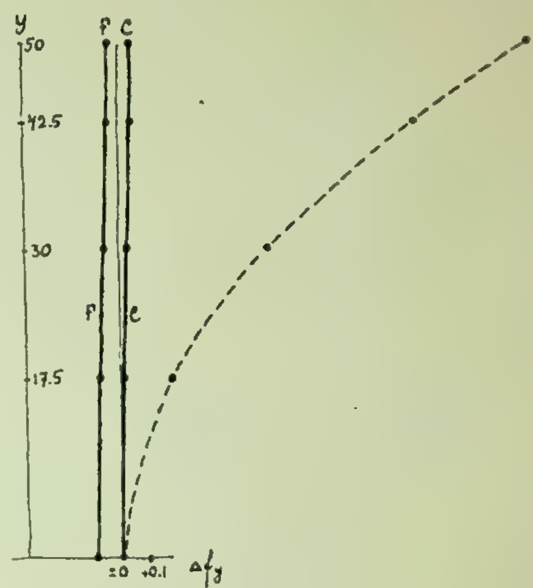


FIG. 3.

be expressed by the formula

$$\frac{f_C - f_F}{f} \cdot 100 \text{ percent} = 0.057A^{1}. \qquad (15)$$

The secondary spectrum is vanishingly small (500–800 times smaller than for equivalent achromatic objectives); therefore its empirical formula will not be given here. Empirical formulae give only approximate values of aberrations, but their accuracy is sufficient to use them to determine the limit diameters and the relative apertures of meniscus systems. Thus, in first-rate visual systems, in which the wave aberrations must not exceed $\frac{1}{4}\lambda$ for monochromatic radiation $\lambda \sim 555$ mμ, the limit diameter $D_{\max\,vis}$ must not exceed

$$D_{\max\,vis} = 1.37V^{4.5}. \qquad (16)$$

In the same way, in photographic systems intended to resolve 30 lines per mm, the limit diameter $D_{\max\,phot}$ must not exceed

$$D_{\max\,phot} = 55.8V^{3.5}. \qquad (17)$$

It is possible, applying the empirical formulae (6)–(10) for optical elements, (13)–(15) for re-

TABLE II. $d : D = 0.1$; glass : $n_D = 1.5163$; $\nu = 64.1$.

	A	1 : 1	1 : 1.5	1 : 2	1 : 2.5	1 : 3	1 : 3.5	1 : 4	1 : 4.5	1 : 5
1	$-R_1 : D$	0.612	0.800	0.967	1.120	1.264	1.399	1.528	1.651	1.770
2	$-R_2 : D$	0.676	0.861	1.027	1.179	1.323	1.458	1.586	1.709	1.828
3	$-R_3 : D$	2.107	3.139	4.165	5.186	6.204	7.219	8.232	9.242	10.25
4	$d_2 : D$	1.110	1.762	2.446	3.155	3.884	4.630	5.391	6.166	6.953
5	$\Delta\eta_{max}$	123″	19.8″	5.44″	1.99″	0.88″	0.44″	0.24″	0.14″	0.09″
6	$\pm\Delta\eta_{C,F}$	7.6″	2.25″	0.95″	0.49″	0.28″	0.18″	0.12″	0.08″	0.06″
7	$\frac{f_C - f_F}{f} \cdot 100\%$	0.057%	0.044%	0.036%	0.031%	0.027%	0.025%	0.023%	0.021%	0.020%
8	$D_{max\ vis}$ (mm)	1.37	8.5	31.0	84.6	192	394	701	1190	1910
9	$D_{max\ phot}$ (mm)	55.8	230	632	1380	2620	>5000			

sidual aberrations, and (16), (17) for limit diameters, to compile Table II, which is useful for all preliminary calculations and for finding a solution.

For instance, with the relative aperture $A = 1 : 3.5$ and optical elements very near to the values given in lines 1, 2, 3, and 4 of the table, the system will be well achromatized and corrected for spherical aberration and for coma. The residual aberrations, according to lines 5 and 6 of the table, may be brought into the form of Fig. 4, where the dotted line corresponds to the aberrations for rays of a certain wave-length λ_0, near to 555 $m\mu$; for these rays $\Delta\eta_{max} = 0.44''$. The spherochromatic aberration for rays C and F leads to a residual chromatism $\Delta\eta_{C,F} \cong 0.18''$. The system calculated in the way described will be a first-rate visual system for diameters not exceeding 394 mm (line 8) and will resolve 30 lines in 1 mm even if its diameter for the same relative aperture (1 : 3.5) exceeds 5 meters, provided that it would be possible to realize such a system practically.

Amateur astronomers are mainly interested in visual instruments of medium size. Table III gives the limit relative apertures for such systems satisfying Rayleigh's condition.

In the last line the distance d_2 is given; that determines the length of the instrument. If the complete correction for coma is not aimed at, a further diminution of the distance d_2 is possible and with it a diminution of the length of the instrument; even in that case the coma will be several times smaller than for an equivalent parabolic mirror. It goes without saying that the relative aperture of the telescope may be reduced as desired compared with the values A_{max} of Table III.

6. RETOUCHED SYSTEMS

Owing to the correction of the spherical aberration and the presence of only a residual spherical aberration in a meniscus system, an insignificant retouching of the spherical mirror B (Fig. 1) is sufficient to straighten the curve λ_0 of Fig. 4 and to eliminate completely the residual spherical aberration for rays λ_0. After such a

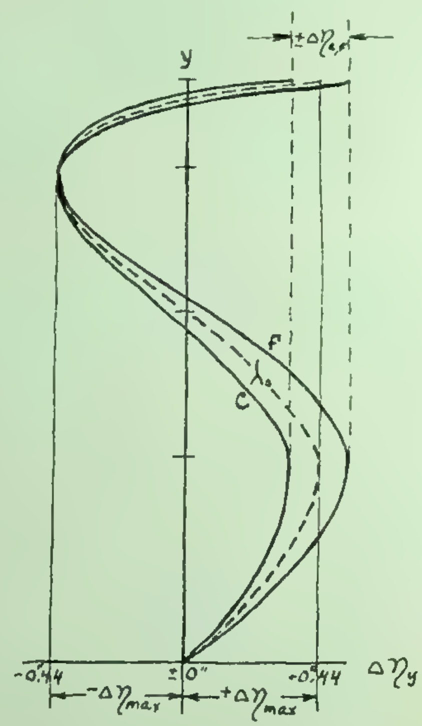


FIG. 4.

retouch the curves of Fig. 4 will acquire the form of curves of Fig. 5 and we obtain a system with only an insignificant residual chromatism.

This easily achieved retouch allows a considerable extension of limits of possible maximal

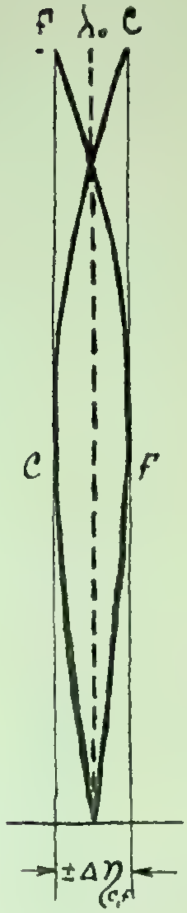


Fig. 5.

diameters and relative apertures of the systems; it may be done not only for the ray λ_0 with infinite theoretical limits, but also for rays C and F, to obtain for them wave aberrations that do not exceed $\frac{1}{4}\lambda$. In other words, we introduce a condition of achromatization as stringent as that used in the derivation of the expression (1) for the achromatic objective a distinction due to the circumstance that while the secondary spectrum of objectives increases strongly with the increase of the spectral interval, the spherochromatic aberration $\Delta\eta_{\lambda_1,\lambda_2}$ of meniscus systems increases both more slowly and according to a more favorable law.

Imposing this more than rigorous condition, the following empirical formula for the limit diameter of first-rate visual systems with a *retouched* mirror may be obtained:

$$D_{\max}=25V^3. \qquad (18)$$

For instance, in a system $D=100$ mm with a retouched mirror the relative aperture may be brought up to $A=1:1.59$ with practically no chromatism in the interval from C to F. In the same way for a system $D=1000$ mm the relative aperture may be brought up to $1:3.42$. In practice these rigorous theoretical limits may be exceeded considerably and even more powerful systems with larger relative apertures may be realized. They will still be considerably more achromatic than the achromatic objective of long focus, the chromatism of which we are tolerating.

Finally, there is a method of eliminating also the spherochromatic aberration of Fig. 5, and thus almost infinitely extending the limits of possible diameters and relative apertures of meniscus systems. This method consists of a combined retouch of the meniscus and of the mirror; however, it will not be described here because it is more difficult and may be of interest only for special instruments of very large relative aperture with higher requirements of achromatization in a wide wave-length region.

Though the retouched systems are more perfect than the non-retouched ones, and allow us to attain greater diameters and relative apertures and to design instruments of a smaller size, meniscus systems with non-retouched optics are so perfect and have a relative aperture so large that without adequate cause it is unreasonable to complicate their construction by retouching, this last being a delicate work, requiring great skill.

Table III.

D (mm)	50	70	100	140	200
$A_{\max}$	1 : 2.22	1 : 2.40	1 : 2.60	1 : 2.80	1 : 3.05
d_2 (mm)	138	211	330	503	786

For production making of optical instruments such a complication of technology is entirely inadmissible.

Therefore in the following also only systems with the simplest spherical non-retouched optics will be considered.

We have pointed out the possibility of transforming any classical system of telescope into a meniscus one. Achieving such a transformation we obtain an essentially new optical system with

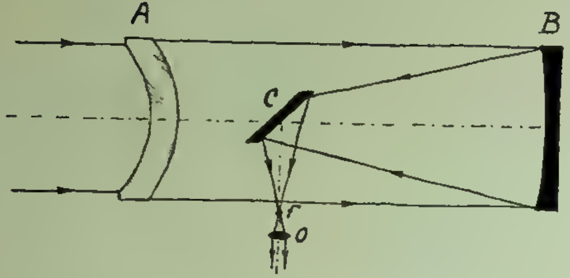


FIG. 6.

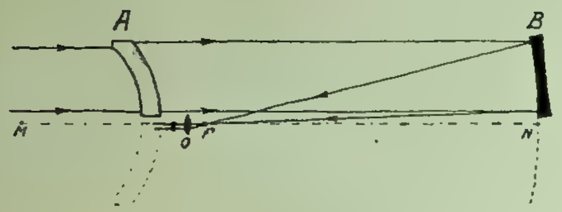


FIG. 7.

new properties; however, to be concise and more descriptive in the following we shall call the systems transformed into meniscus ones by the names of their initial authors.

7. NEWTONIAN AND HERSCHELIAN SYSTEMS

A meniscus-concave mirror system may be used to take photographs directly in the focus F (Fig. 1), but it is not suitable for visual observations. The addition of a diagonal mirror C (Fig. 6) transforms it into a Newtonian system; all the results obtained for the meniscus-concave mirror system hold in full extent also for the transformed system. For diminishing the screening by the diagonal mirror in systems with large relative apertures and for diminishing the relative aperture, which is advantageous to make possible the use of less perfect eyepieces, the mirror C may be combined with a scattering system, similar to a Barlow lens.

If from a meniscus-concave mirror system a part is cut out that is somewhat less than a half of its diameter, a transformed Herschelian system is obtained (Fig. 7) with the eyepiece O directed along the axis MN of the system and with an eccentric exit pupil.

For such a system all the considerations for the meniscus-concave mirror systems hold with the difference that the part cut out from the system will have somewhat smaller aberrations than the entire system.

The decentered meniscus A and also its inclination and the inclination of the mirror B with respect to the tube axis cause some complications in the making of optics and mechanics. However, these complications are not so great as to cause us to give up the possibility of a very advantageous instrument showing maximal light transmission, having a filled exit pupil, and free from screening elements and diffraction hindrances. It goes without saying that in contrast with classical Newtonian and Herschelian systems the meniscus systems considered may be made free from coma. The computations show also that the astigmatism may be practically eliminated; the remaining aberration is the field curvature, but the methods to correct it are known to opticians (the introduction of correction lenses in front of the plate) and to astronomers (photographs taken on convex spherical films).

However, for visual observations neither the astigmatism nor the field curvature is of importance. The Herschelian system (Fig. 7) is of a great interest for collimators and other laboratory and control instruments in which the screening of the beam by the secondary mirror is undesirable. With the help of a small prism of total reflection it is possible in a Herschelian system to break the axis at 90° as in a Newtonian system, with all the advantages of a wholly unscreened beam remaining.

8. GREGORIAN AND CASSEGRAINIAN SYSTEMS

Figures 8 and 9 give an idea of such systems. It is to be noted that in many cases it is possible to do without a special small secondary mirror; its role may be filled by the aluminized central part C–C of the inner surface of the meniscus (Fig. 8); in a more general case, as, for instance, that shown by Fig. 9, the secondary mirror C–C may be glued to (or even polished on) the inner surface of the meniscus.

Such systems have already been computed, constructed, tested, and quite satisfactory results were obtained.

Figure 10 shows a photograph of a telescope $D=100$ mm, made according to the scheme of Fig. 8 (Gregorian system), giving direct images without any additional optics.

In a Cassegrainian as well as in a Gregorian system, a real diaphragm d–d must be placed

behind the eyepiece O to cut out unwanted light rays.

If it is desired to obtain a direct image in a Cassegrainian (Fig. 9), an astronomical eyepiece O has to be used instead of a terrestrial one; the use of the latter allows us to shift the diaphragm d–d inside the eyepiece and to place the eye conveniently without a material obstacle in the form of the diaphragm d–d.

If a part is cut out from a Gregorian or from a Cassegrainian in a way similar to that used in the transition from a Newtonian to a Herschelian, an instrument is obtained similar to the "brahite."

Figures 11 and 12 show the schemes of two versions of the transformed brahites, corresponding to the two initial systems from which they are cut out.

Gregorian and Cassegrainian systems are typical representatives of teleobjectives, hence they afford the realization of instruments of a small size with a great focal length. For stellar photography an instrument of great interest is that obtained by following the scheme of Fig. 13, in which projection optics of moderate size shift the image from F to F' and at the same time allow us to place a real diaphragm d–d, cutting off the unwanted rays, veiling the plate. The image in the focus F' may be to a high degree achromatic, aplanatic, and astigmatic; the field curvature may be handled as mentioned above in one way or another.

Fig. 8.

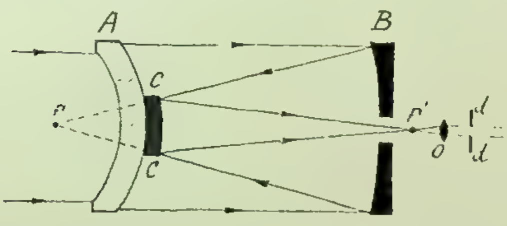


Fig. 9.

Fig. 10.

Uviol optics with clarified surfaces and the aluminizing of the mirrors B and C–C allow us to obtain a considerable increase of light transmission for actinic rays.

9. SCHMIDT SYSTEMS

The Schmidt system, though remarkable in its idea, has an essential practical defect: The asphericity of its plates is about four times greater than the asphericity of a parabolic mirror of the same diameter and of the same relative power as the Schmidt system. Therefore, notwithstanding the smaller accuracy required, the making of such aspherical plates is often at the limit or outside the limit of the skill of the technician. Schmidt systems of great diameter and with great relative aperture either have an image quality diminished by the optician, or require

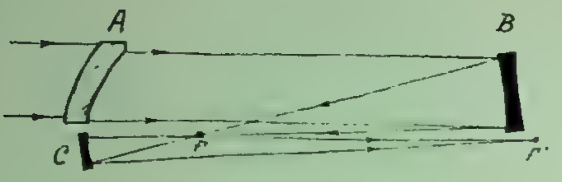


FIG. 11.

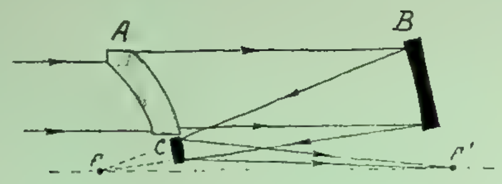


FIG. 12.

considerable time and expense and a high skill for their construction. Besides, in a Schmidt system, the distance between the end optical elements (plate and mirror) that determines the length of the instrument is twice its focal length.

In a transformed Schmidt meniscus system the role of an aspherical plate is played by the meniscus A (Fig. 14).

It is possible to compute the system in such a way that the centers of curvature of both surfaces of the meniscus and of the mirror are near to coincidence in a certain point O. The aperture diaphragm D–D has to be placed in the plane of that point and photographs of the image in the focus F have to be taken on a convex film.

In a special case, giving up a perfect achromatization though without going outside the limits allowable, it is possible to replace the achromatic meniscus A by a meniscus with concentric surfaces and to attain a rigorous coincidence of centers of curvature of all the three surfaces of the system, as well as of the focal plane, in the point O. In that case, even for an unlimited field, not only the coma and the astigmatism of the third order will be eliminated, but also their higher orders, i.e., the residual aberrations, a thing altogether impossible in an ordinary Schmidt system.

Preliminary computations show that the transformed Schmidt system with spherical nonretouched optics and a diameter $D=200$ mm may have a relative aperture of 1 : 1.4. For a diameter $D=1$ m such a system may have a relative aperture of about 1 : 2.3, may resolve 30 lines in 1 mm, may have a field of vision greater than ten degrees, and may be very simple and cheap to make.

The length of the rigid tube connecting the meniscus A and the mirror B of Fig. 14 turns out to be shorter than twice the focal length of the system; and though the distance from the diaphragm D–D to the mirror B is not less than the length of the usual Schmidt system it must be kept in mind that this diaphragm may be fastened in a light additional tube.

The Schmidt system is a typical example of wide-angled systems with great relative aperture. It must be remembered that the simplest meniscus-concave mirror system may have a relative aperture as great and may have a sufficiently wide angle while gaining considerable advantages in regard to size.

10. THE MENISCUS IN CONVERGENT RAYS

In all systems considered above the meniscus has been placed in parallel rays; hence the diameter of the meniscus corresponds to the diameters of the active apertures of the systems.[4] The reader cannot avoid feeling that in such a construction the practical limit of the diameter of the system is imposed not by the aberrations, but by the difficulty of obtaining sufficiently great masses of optically homogeneous glass for the meniscus. The present state of optical glassmaking allows the attainment of a meniscus of a diameter of 1 m but hardly more. Therefore the limit diameter of a meniscus system of any of the types considered above probably is close to 1 meter, and in any case does

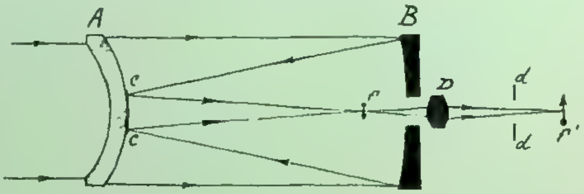


FIG. 13.

not exceed 1.5 meters, whereas astronomers may be interested in instruments of even greater aperture.

[4] In a Schmidt system with a considerable field and with the condition of absence of vignetting in inclined rays, the diameter of the meniscus may be even somewhat greater than the diameter of the active aperture.

On the other hand, meniscus systems transparent to the far ultraviolet require for the meniscus the use of optical fused quartz or of an analogous substance, i.e., of materials that are very expensive and cannot be obtained in great masses.

Therefore it should be advantageous to diminish the mass of the meniscus material and thus to increase the diameter of the aperture of the meniscus system.

Giving up the valuable property of the systems described above—their air-tightness—it is possible to suggest systems with menisci in convergent rays, that is, equivalent to a considerable decrease of the diameter of the meniscus and of the mass of its substance.

Let us consider here only one such scheme, which by analogy may be called a *transformed Rossian system* (Fig. 15). The concave spherical mirror B gives an image at the focus F; the convergent beam with aberrations is intercepted by the meniscus A and the final image is obtained at the focus F.

The problem for the calculator consists of the choice of such a meniscus, which will correct the spherical aberration and the coma of the mirror B without introducing a noticeable chromatism.

As an example, the results of the trigonometrical computation of the following system are given in Table IV:

$D=100,\quad A=1:3.1,$
$R_1=+600,$
$$d_1=200, \tag{19}$$

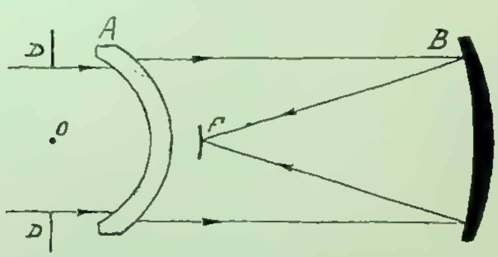


FIG. 14.

$R_2=-52,$
$$d_2=10\ (n_D=1.5163,\ \nu=64.1).$$
$R_3=-53.65,$

Figure 16 shows the residual angular aberration $\Delta\eta_y$ for three colors: C, λ_0, and F; it is seen that for monochromatic light (λ_0) the angular

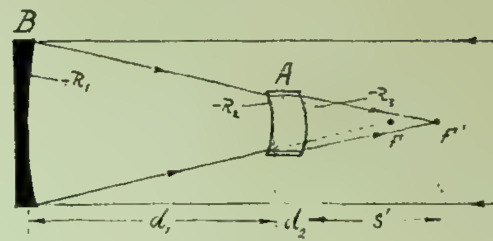


FIG. 15.

aberrations of the system do not exceed ± 1 arc second. This system is not corrected for coma; however, its coma is only a half of the coma of an equivalent parabolic mirror. For a complete correction of the coma the thickness and the curvature of the meniscus ought to be somewhat increased; that should make the residual spherical aberration somewhat worse.

As the diameter of the meniscus in that system is about three times smaller than the diameter of the active aperture D and as formerly $d_2 : D=0.1$, this system, compared with those considered

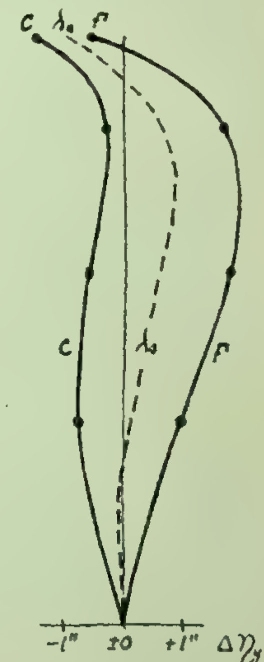


FIG. 16.

above, requires about a nine times smaller mass of glass for the meniscus.

It must be pointed out that this system is far from an optimal solution of the problem and only represents an example of a possible solution.

TABLE IV. $S' = 103.160$.

Zone y	17.5		30		42.5		50	
Ray	C	F	C	F	C	F	C	F
$\Delta S_y'$	−0.021	+0.026	−0.010	+0.029	−0.003	+0.020	−0.014	−0.006
$\Delta\eta_y$	−0.8″	+1.0″	−0.6″	+1.9″	−0.3″	+1.8″	−1.5″	−0.7″
f_y	309.817	309.643	310.138	309.942	310.558	310.323	310.803	310.530

We may, for instance, increase the thickness of the meniscus A considerably and at the same time decrease its diameter, placing the meniscus nearer to the focus F. Such a thick meniscus of a moderate diameter may be transformed into a prism of total reflection with concave and convex catheta surfaces, producing a peculiar Newtonian telescope (Fig. 17), in which the element A simultaneously acts as a diagonal mirror and as a correcting meniscus placed in convergent rays. In this system all optics may be spherical and the system may prove to be not only achromatic and stigmatic, but also aplanatic, the last being not indispensable for visual amateur telescopes.

11. AMATEUR TELESCOPES

The professional optician, taking the idea of meniscus systems, may compute and reproduce any scheme, whether mentioned or not.

But the amateur, often having no knowledge of the technique of computations, needs more detailed recommendations concerning the types and the optical elements of the instrument.

It is impossible to give exhaustive directions regarding optical elements, because of lack of knowledge of the constants of the glass which the amateur will use. However, it must be kept in mind that the aberration properties of meniscus systems show a relatively feeble dependence on the index of refraction of glass; that is another advantage of meniscus systems over objectives. Hence, to avoid optical computations, the amateur may buy crown glass for the meniscus, having an index of refraction sufficiently near if not equal to $n_D = 1.5163$. After this he may use the empirical formulae given above and Table II in the following manner.

Let us suppose that the amateur wants to make a Newtonian meniscus telescope with an aperture $D = 200$ mm. Formula (16) as well as Table II gives $A_{max} = 1 : 3.03$ for the limit relative aperture of such a telescope. To reach this maximum aperture an exact trigonometrical computation and an exact knowledge of optical constants of glass are needed. However, if a somewhat smaller relative aperture is taken, e.g., $A = 1 : 3.5$, it is possible to compute a telescope of passable quality with the help of empirical formulae. Applying formulae (6)–(10) or the data of Table II we find the following optical elements of the system:

$$\begin{aligned} &D = 200, && R_1 = -279.8, && d_1 = 20, \\ &A = 1 : 3.5, && R_2 = -291.6, && d_2 = 926. \\ &f = 700, && R_3 = -1444, && \end{aligned} \tag{20}$$

With $d_2 = 926$ the system will not only be stigmatic, but also well corrected for coma. With $A = 1 : 3.5$ only achromatic eyepieces of high quality need to be used.

Let us now suppose that the amateur wishes to design a Herschelian system with an aperture of 200 mm, cutting it out, as a part, from the meniscus-concave mirror system with a diameter $D = 430$ mm and using a small prism of total reflection to break the axis at 90° near to the focus.[5]

With $D = 430$ mm we have $A_{max} = 1 : 3.59$. If a somewhat smaller relative aperture is taken,

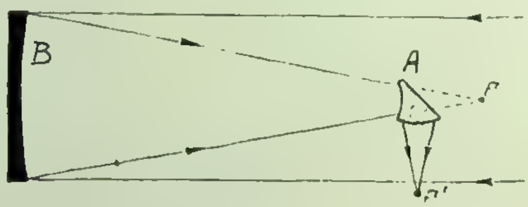


FIG. 17.

e.g., $A = 1 : 4$, we obtain the optical elements, first of the main system and secondly of the Herschelian, cut out from it. See Table V. With

[5] It is particularly advantageous to use for such a telescope a mounting of the "Springfield" type.

TABLE V.

Main system		Herschelian system	
$D=430$		$D=200$	
$A=1:4$		$A=1:8.6$	
$f=1720$		$f=1720$	
$R_1=-657$	$d_1=43(n_D=1.516)$	$R_1=-657$	$d_1=43(n_D=1.516)$
	$d_2=2318$		$d_2=2318$
$R_2=-682$		$R_2=-682$	
$R_3=-3540$		$R_3=-3540$	

$d_2=2318$ the system is not only stigmatic but also corrected for coma.

It must be remembered that in "meniscus-concave mirror" systems the aperture D is determined by the diameter of the first surface of the meniscus (Fig. 6); within the meniscus the rays form a divergent beam, hence the second surface has a diameter D' somewhat greater than D; finally the diameter of the concave mirror D'' is somewhat greater than D'. It should be possible to give computation formulae to find the active apertures; however, these apertures may also be determined by drawing the path of the rays. Their exact values are usually obtained from a trigonometrical computation.

For complex systems (Gregorian, Cassegrainian, etc.) a trigonometrical computation is almost unavoidable. For amateurs familiar with the art of calculation, a Gregorian system may be recommended having an aperture of about 200 mm and a set of eyepieces of the simplest type but provided with ocular caps with specially calculated holes. Such a telescope will act simultaneously as a powerful terrestrial instrument (erect images).

A Cassegrainian of maximum aperture will always be shorter than the Gregorian maximum aperture of the same diameter, hence a Cassegrainian telescope is more advantageous than any other system in respect to size. To cut out possible stray illumination of the field the ocular caps in a Cassegrainian likewise ought to be made according to a special calculation. Only instead of an astronomical eyepiece a special terrestrial one may be used, as has been mentioned above.

Finally, a system shown in Fig. 17 may be recommended to the attention of amateurs; it has, however, all the defects of the systems with open tubes.

Having at his disposal an autocollimating plane mirror and being familiar with the knife edge testing method, the amateur may try to obtain still shorter and more compact telescopes (e.g., Cassegrainian) with a large retouched mirror. The retouching of the mirror produces also a considerable improvement of the quality of the telescope, if the amateur, not being familiar with the art of calculation and not being in possession of glass of the kind similar to that recommended, has failed in designing the optics.

The scheme of the arrangement for retouching is shown by Fig. 18, where A is a meniscus with an aluminized central part C–C, B is a retouched mirror, D is an autocollimating plane mirror, E is the luminous point, and F is a Foucault knife.[6] In this scheme the mirror B has to be retouched until a plane knife edge relief is obtained, when $\Delta\eta_y=0$. It is desirable to point out that a Gregorian (or Cassegrainian) meniscus system of

FIG. 18.

a diameter $\sim$100 mm, with two magnifications ($W_1\sim 35X$ and $W_2\sim 100X$) and with rationally designed mechanics, forms an excellent and cheap school telescope. Such a telescope will have a price which will make it accessible not only for urban but for rural schools as well and will allow the teacher to arouse the interest of many young people in astronomy.

12. CONCLUSION

In the present short review only a small number of the problems worked out by the author have been considered. In the postwar reconstruction period great possibilities will be opened for the attainment of large meniscus systems to equip our observatories.

As a large angled astrograph with a large relative aperture we may recommend a Schmidt

[6] The distance from E to F must not exceed several millimeters. In knife edge instruments of the author this distance may be decreased in a part of a millimeter.

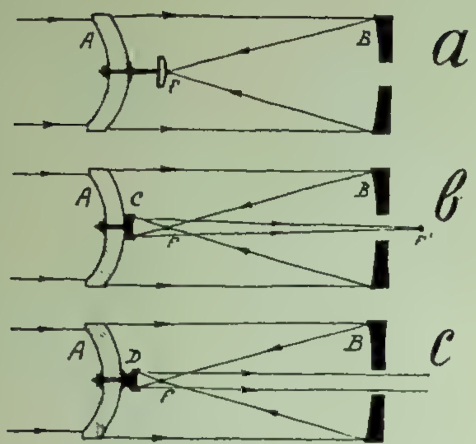


FIG. 19.

meniscus system with a diameter up to 1 meter and with a relative aperture up to 1 : 2.

Systems of smaller size, but with a greater relative aperture, having objective prisms as well, will act as powerful wide-angled spectrographs of very large relative aperture and no slit.

To obtain different observations with the help of the same instrument the following system of the combined type may be recommended: A meniscus A and a retouched mirror B (Fig. 19a) form a system of maximum relative aperture for taking photographs at the focus F. In this case the system must be free from chromatism, spherical aberration, coma, and astigmatism; the last allows us to count on large fields of vision. Placing a concave elliptic mirror C (Fig. 19b) instead of the plate holder with convex film, we transform the telescope into a Gregorian of long focus and the visual observations are made or photographs taken at the focus F', where a correction for chromatic and spherical aberration takes place, but where a perfect correction for coma and still less for astigmatism is not indispensable. In the focus F' may be placed the slit of a spectrograph. Finally, placing a parabolic mirror (Fig. 19c) instead of an elliptic one C, we obtain an analog of one of the Mersennian systems, serving to send parallel beams into a spectrograph without slit.

The present state of glassmaking affords a full possibility of making such a telescope with a diameter of about 1 meter.

At last it is possible to obtain an aplanatic telescope of a diameter of about 3 meters, according to the scheme of Fig. 15.

Meniscus systems may be used in different astronomical instruments in the form of finders, cameras of spectrographs, finders of comets, cameras for taking photographs and spectrograms of meteors, etc. Solely for accurate astrometrical instruments the meniscus systems are not suitable, as, in that case, systems are inadmissible that contain more than one mirror element.

Influence of Pneumatic Pressure on the Photographic Sensitivity

CHOONG SHIN-PIAW
Institute of Physics, National Academy of Peiping, Kunming, China

(Received February 23, 1944)

The effect of pneumatic pressure on photographic sensitivity for yellow and violet light is investigated. Three different gases, namely, N_2, CO_2, and H_2, were used in the experiment to produce the pressure. In the case of N_2 and CO_2, all the films show a lower sensitivity when they are subjected to pneumatic pressure during exposure. On the other hand, the application of pneumatic pressure before exposure increases the sensitivity of the films. No appreciable modification of sensitivity is observed when pressure is applied after the exposure. In the case of hydrogen, the result is somewhat different: All the films show generally higher sensitivity when they are subjected to pressure either during or before exposure. In all cases, the effect of pneumatic pressure on photographic sensitivity, as interpreted by the mean value of the variation of density from the portion of normal exposure on the characteristic curves is found to be independent of the wave-length of light. The abnormal behavior of hydrogen on photographic sensitivity is considered to be due to the chemical activity of the gas on the silver bromide of the emulsion. Certain differences between the effects produced by *pneumatic* and *mechanical pressures* on photographic sensitivity are pointed out and discussed.

INTRODUCTION

THE application of mechanical pressure[1] to a photographic emulsion during exposure prevents the formation of latent image to a great extent and thus reduces its sensitivity. This effect of pressure, first observed by F. E. Poindexter[2] in 1931, was subsequently much studied by Ny Tsi-Ze[3] and his associates, and more recently by Reardon.[4] To characterize this desensitizing effect, Ny and Tsien had introduced a quantity E/E_e, E_e being the illumination which would produce on an emulsion without being subjected to any pressure, the same optical density as that produced by illumination E on the same emulsion under a pressure p with equal times of exposure. According to Ny and Tsien, the quantity E/E_e is an increasing function of pressure p only, and for p greater than a certain value, the ratio E/E_e is a linear function of pressure. The same workers had also established that the desensitizing effect, produced by applying mechanical pressure during exposure, decreased generally with the wave-length of light. The dependence of the effect of mechanical pressure on the wave-length is indeed very great; for example, for the yellow light a pressure of some eighty kilograms per square centimeter is sufficient to produce a marked effect, while for the ultraviolet radiation 3131A, a pressure of more than one thousand kilograms per square centimeter will render the effect only just observable.

All the investigators mentioned above employed the same kind of pressure in their experiments, namely, the mechanical pressure produced by a lever. This kind of pressure though easy to produce is not quite suitable for the following reasons. Firstly, owing to the variation of their size, the sensitive grains of an emulsion under pressure do not actually receive the same amount of pressure. As had been pointed out by Poindexter, the unevenness in pressure is most clearly manifested in emulsion of large grains. Secondly, the application of mechanical pressure modifies the surface state of an emulsion. As a matter of fact, the surface is greatly smoothed, resulting in a marked change of reflective power. Apparently, if the pneumatic pressure is employed, instead of the mechanical one, these difficulties will be removed. It is with this new kind of pressure that the writer has carried out a systematic study of the pressure effect on photographic sensitivity. Some preliminary results have already been indicated in a previous note.[5]

[1] The effect caused by mechanical pressure will be specified in this paper.
[2] F. E. Poindexter, J. Opt. Soc. Am. **21**, 59 (1931).
[3] Ny Tsi-Ze and Tsien Ling-Chao, Comptes rendus **194**, 1644 (1932); Sci. Ind. Phot. [2], **4**, 1 (1933); Ny Tsi-Ze, IX Congrès International de Photographie (1936), p. 83; Ny Tsi-Ze and Lu Ta-Yuan, J. Opt. Soc. Am. **26**, 26 (1936).
[4] A. J. Reardon, J. Opt. Soc. Am. **29**, 427 (1939).
[5] Choong Shin-Piaw, J. Opt. Soc. Am. **31**, 186 (1941).

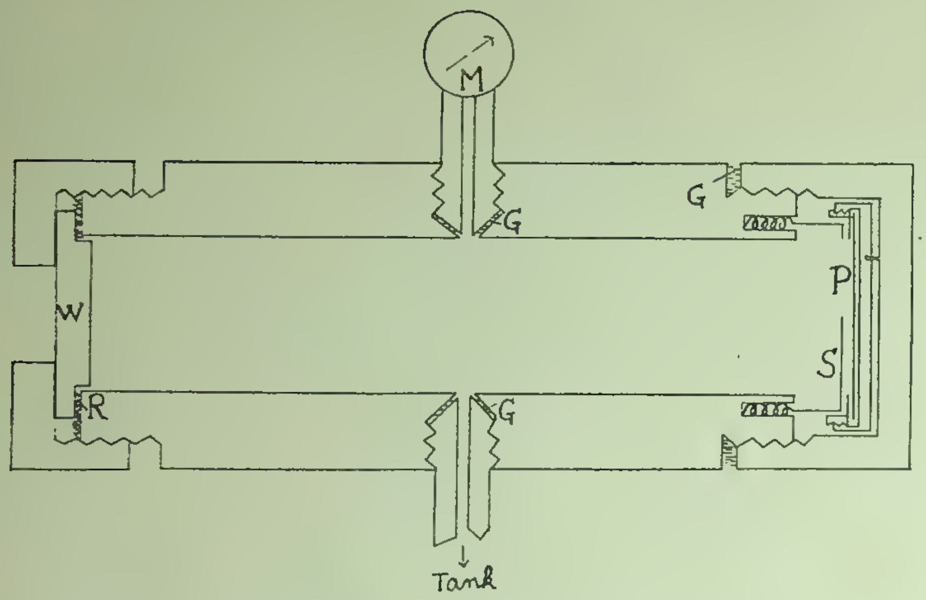


FIG. 1.

APPARATUS AND EXPERIMENTAL PROCEDURE

The photographic chamber consists essentially of a thick brass tube with a thick quartz window on one end and a circular film or plate holder on the other. It is schematically shown in Fig. 1. The chamber 15 cm long, 5-cm inner diameter, has at its middle part two side tubes; the one leads to the compressed gas tank, the other to the manometer M. The quartz window W and the plate holder P are each held by a brass frame threaded on to the main tube. A rubber ring R prevents the leakage of gas along the rim of the quartz window and lead rings are employed in the gaskets G. A spring-supported semicircular screen placed in front of the plate holder permits one to make two separate exposures on the same film. For minimizing the scattering of light, the inner wall of the photographic chamber is blackened with copper oxide.

Two kinds of light, the violet (4040–4300A) and the yellow (5720–5920A) were used;. they were obtained by inserting, respectively, the solution filters of potassium permanganate and cobalt sulfate in front of a small electric lamp.

The pressures employed in the experiments were furnished by the compressed gases: N_2, H_2, and CO_2. The highest attainable pressure of the first two gases was about 150 kg/cm^2 and of the last about 27 kg/cm^2.

Three kinds of films were used, namely, Eastman Ortho-X, Eastman Super-Speed Portrait, and Agfa Isochrome films of which the last two have been used by previous workers[6] in their studies with mechanical pressures. The exposed films were developed in the Eastman high contrast developer (formula D-19) under practically identical conditions so that the effect on the different films can be compared.

The electric lamp, provided with a case and a socket, is situated at about 120 cm from the photographic chamber. In actual practice it was necessary to demount the photographic chamber for replacing films. This operation, though rather tedious, would not disturb the optical settings, for with the guide of a spirit level, the chamber after demounting can be remounted in its original position.

On each circular piece of film two exposures of equal time were generally made, one at normal atmospheric pressure, and the other at given high pressure either during or before exposure. For a given light and a given pressure, seven circular pieces of film of different exposing time were usually made. The measurements of the densities of the parts under exposure with a Moll microphotometer permit one to plot two density-log time curves, one at normal pressure while the other at the given high pressure.

[6] Ny Tsi-Ze and Tsien-Ling-Chao, Chinese J. Phys. 1, 66 (1934).

RESULTS

The results are classified into two groups, according to the nature of the gases.

(1) The case of nitrogen and carbon dioxide: All the three kinds of films showed a lower sensitivity when they were subjected to pneumatic pressure during exposure. The desensitizing effect, as interpreted by the mean value of the variation of density from the portion of normal exposure on the characteristic curve, was found to increase with the pressure applied, and vary with the kind of gas and also the emulsion employed. However, within the limit of experimental error, the effect appeared to be independent of the wave-length of the exposing light. On the other hand, the application of pneumatic pressure before exposure increased the sensitivity of the films. The sensitizing effect was found to be practically independent of the wave-length of light and vary very little with the magnitude of pressure previously applied. No appreciable modification of sensitivity could be observed when pressure was applied after the exposure.

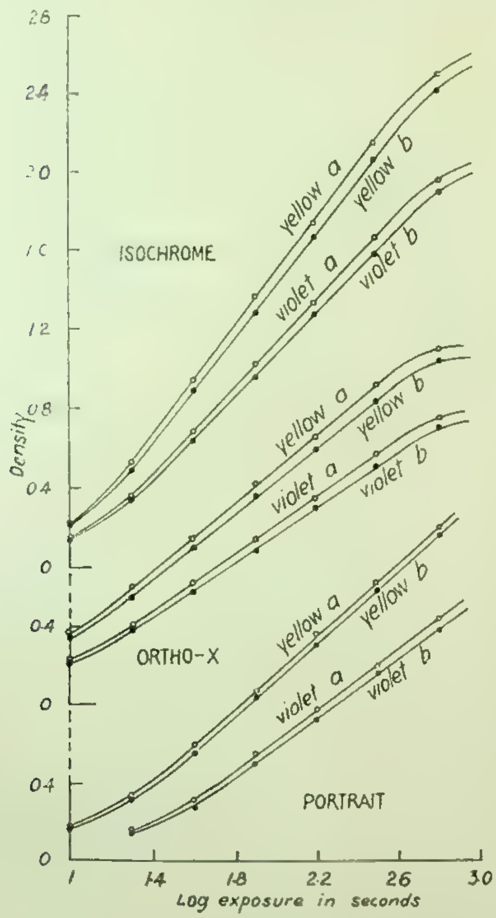


FIG. 2. The desensitizing effect on Isochrome, Ortho-X, and Portrait films produced by the application of 150 kg/cm² of pneumatic pressure of N_2 during exposure.

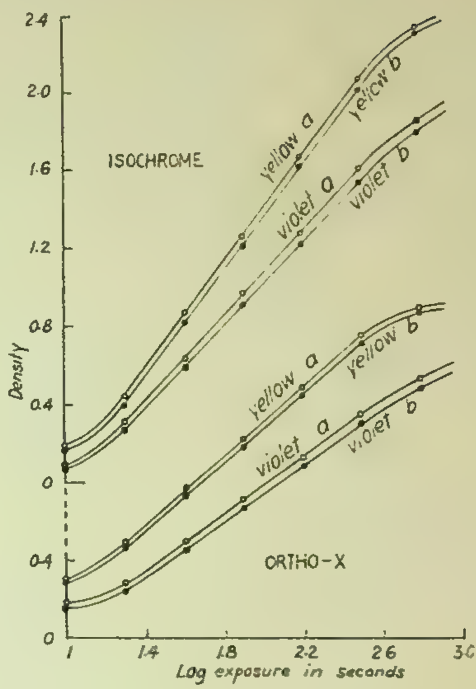


FIG. 3. The desensitizing effect on Isochrome and Ortho-X films, produced by the application of 25 kg/cm² of pneumatic pressure of CO_2 during exposure.

The desensitizing effect in the case of nitrogen for different light on Agfa Isochrome, Eastman Ortho-X, and Eastman Portrait films is represented in Fig. 2 and that of carbon dioxide on Isochrome and Ortho-X films in Fig. 3. The curves marked *a* represent the results of the films exposed at normal atmospheric pressure and those marked *b* represent the results of the films to which a pressure of 150 kg/cm² (in case of N_2) or of 25 kg/cm² (in case of CO_2) was applied during exposure.

The curves for the case in which the pressure was applied before exposure are similar in appearance to those of Fig. 2 and Fig. 3 except that the positions of curves *a* and *b* are interchanged.

(2) The case of hydrogen: The Ortho-X and Portrait films showed always higher sensitivity when they were subjected to the pressure of hydrogen either during or before exposure, while

the pressure effect on Agfa Isochrome film was somewhat different. For Agfa Isochrome film, the sensitivity became higher when exposed to violet light, but when yellow light was used no change of sensitivity can be observed. All the films did not show any alternation of sensitivity when they were subjected to the pressure of hydrogen after exposure. Figure 4 represents the sensitizing effect caused by the pressure of hydrogen. The curves *a* are for films exposed at normal atmospheric pressure and those marked *b* for films to which a pressure of 150 kg/cm² was applied during exposure.

DISCUSSION

The abnormal behavior of hydrogen on photographic sensitivity may be closely related with the chemical activity of the gas. In fact, silver bromide is known to be reducible by the electrolytic hydrogen (in nascent state) at ordinary temperature and silver chloride[7] is also known to be reducible by hydrogen at high pressure. It seems fairly probable that the silver bromide of an emulsion is also reducible under a high pressure of hydrogen. The fact that there is no modification of sensitivity when the pressure is applied after the exposure tends to show that light energy is indispensable for the provocation of the eventual chemical reaction in question. Perhaps certain impurity may also have great influence on this chemical reaction as the peculiar phenomenon in Agfa Isochrome film tends to indicate.

Owing to molecular scattering, a certain amount of light will be lost, as it travels through the column of compressed gas in the photographic chamber. It is necessary then, to see to what extent this loss will affect the observed experimental data. The transparency coefficient $\alpha = I/I_0$ of a column of gas may be calculated from the expression:

$$\log \alpha = -\frac{32}{3}\pi^3 \frac{(\mu-1)^2}{n} \frac{1}{\lambda^4} ph,$$

where μ is the refractive index of the gas, n the number of molecules per cubic centimeter at normal conditions, λ the wave-length of light, p the pressure in atmospheric unit, and h the length of the photographic chamber. Knowing the transparency coefficient α and the approximate value of the contrast γ of an emulsion, it is possible to deduce the variation of density ΔD due to molecular scattering. The values of ΔD for nitrogen at 150 kg/cm², carbon dioxide at 25 kg/cm², and hydrogen at 150 kg/cm² are, respectively, 0.0006, 0.0004, and 0.0000. The mean variation of density as indicated in Figs.

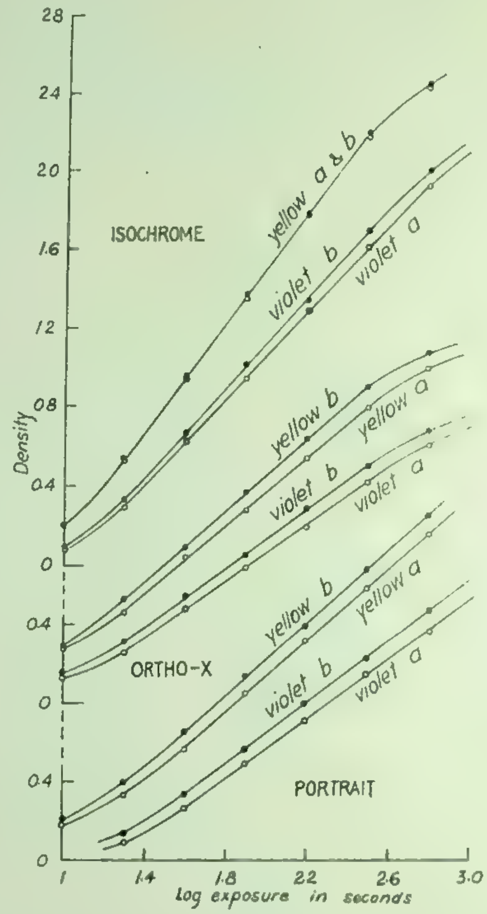


FIG. 4. The sensitizing effect on Isochrome, Ortho-X, and Portrait films produced by the application of 150 kg/cm² of pressure of H_2 during exposure.

2–4 is of the order of 0.07. Hence the molecular scattering effect is practically negligible.

According to Ny and his co-workers, the application of mechanical pressure during exposure will increase the inertia of an emulsion, but leave the contrast unchanged. In the case of pneumatic pressure, however, both the inertia and the contrast are modified by the pressure, as is shown in Figs. 2 and 3. It is rather remark-

[7] D. Tommasi, Ber. 11, 1249 (1878).

able that the desensitizing effect caused by pneumatic pressure is practically independent of wave-length, in contradistinction to the effect caused by mechanical pressure where the yellow has an effect very much greater than the violet. The fact that Agfa Isochrome film shows a desensitizing effect caused by pneumatic pressure greater than Eastman Portrait film (see Fig. 2) is likewise remarkable, as the reverse is true in the case of mechanical pressure. There are obviously certain differences between the effects produced by pneumatic and mechanical pressures on photographic sensitivity.

It has already been stated in the introductory paragraph that the sensitive particles of different sizes of an emulsion do not receive the same amount of pressure when they are macroscopically subjected to a given mechanical pressure. A brief consideration will show that besides this character of non-uniformity in pressure, there is still one particularity which is also uncommon for the case of pneumatic pressure. A particle, subjected to a mechanical pressure, will generally be contracted in the direction of pressure and dilated in directions normal to it. Thus, the particle will actually undergo both contracting and tensile strains. As the effect of pressure is concerned, the case of mechanical pressure appears to be more complicated than that of the pneumatic one. The inconsistency in results of the photographic effect between the mechanical and pneumatic pressures may probably be due to the difference in mechanism of the two pressures on sensitive particles of an emulsion.

In conclusion, the writer wishes to thank Dr. Ny Tsi-Ze, Director of the Institute of Physics, National Academy of Peiping, for his interest and encouragement in the work.

An Easily Cleaned Measure for Powders Used in Spectrography*

CYRUS FELDMAN
Geological Survey, Washington, D. C.
(Received February 1, 1944)

THE chief disadvantage in the use of a scoop for powders used in spectrochemistry has been the difficulty of cleaning particles out of the inside corners of small cavities. The measure *A* here illustrated (see Fig. 1) has no inside corners. It is made of high carbon steel, tempered in oil. The walls of the hole define a frustum of a 60° cone; the lower opening, which is 1.5 mm in diameter, permits efficient emptying and cleaning. For filling, it is placed on a flat surface and the powder is tamped in gently and levelled off with a straightedge. When ground to 270 mesh or finer, rock powders or mixtures of quartz and artificial oxides with as much as 70 percent of quartz are retained by the measure. To insure such retention, the hole in the bottom of *A* should not exceed 1.5 mm in diameter. The measure used by the writer is 1×16 cm, with a hole at each end. It is 0.7 mm thick at one end; the thickness is stepped up to 2.5 mm at the other end to provide for the second hole, which will accommodate a larger charge with no increase in the diameter of the lower opening. If difficulty is experienced with cohesion, the powder may be dampened, provided the water is driven off with a Bunsen flame before the arc is struck.

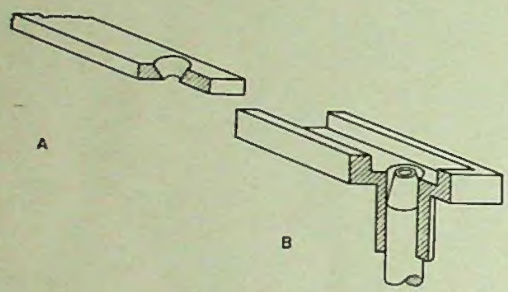


FIG. 1. *A*—the measure. *B*—the centering guide in place over the electrode.

The charge is transferred to the electrode by inserting the measure into the centering guide *B*, shown in place on the electrode, and pushing the powder down with a small brush. The electrodes used in this laboratory are turned from $\frac{1}{4}''$ diameter spectrographic carbon rods. The cavities are $35\frac{1}{2}°$ cones, drilled to provide openings with a diameter of 1.8 or 2.5 mm, and bevelled as illustrated.

The measure may easily be cleaned by drawing a pipe cleaner through the hole.

The author wishes to express his thanks to Mr. A. H. Frazier and Mr. A. C. Wolf of the Geological Survey for their valuable suggestions, and to Mr. Wolf for making the measure.

* Published by permission of the Director, Geological Survey, U. S. Department of the Interior.

THE AMERICAN INSTITUTE OF PHYSICS

INCORPORATED

ADMINISTRATIVE STAFF

HENRY A. BARTON, *Director*

MADELINE M. MITCHELL, *Publications Manager*

JOHN T. TATE, *Adviser on Publications*

Office: 57 East 55 Street, New York 22, New York

JOURNALS

THE PHYSICAL REVIEW
REVIEWS OF MODERN PHYSICS
JOHN T. TATE, *Editor (on leave)*
J. W. BUCHTA, *Acting Editor*

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA
GEORGE R. HARRISON, *Editor*

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA
FLOYD A. FIRESTONE, *Editor*

AMERICAN JOURNAL OF PHYSICS
DUANE ROLLER, *Editor*
JOSEPH D. ELDER, *Assistant Editor*

THE REVIEW OF SCIENTIFIC INSTRUMENTS
GAYLORD P. HARNWELL, *Editor*

THE JOURNAL OF CHEMICAL PHYSICS
JOSEPH E. MAYER, *Editor*

JOURNAL OF APPLIED PHYSICS
ELMER HUTCHISSON, *Editor*

The American Institute of Physics is a membership corporation of which five societies are the founding members. These five societies are the principal national societies devoted to pure and applied physics in America. Their names, and those of their representatives who form the Governing Board of the Institute, are given above. The principal work of the Institute is the publication of journals sponsored by the societies or by itself. The names and the editors of the journals are also listed above.

The broad purpose of the Institute is to represent, in all matters of wide or common interest, the five thousand or more members and subscribers associated with the Founder Societies and the journals. It aims to advance the science and profession of physics, and to promote cooperation between pure research, the applied sciences, and the industries. To achieve these aims, it is greatly in need of an endowment and is empowered by law to administer grants and funds of any kind.

LANCASTER PRESS, INC., LANCASTER, PA.